Deterministic Counting of Graph Colourings Using Sequences of Subgraphs.

Charilaos Efthymiou *

University of Warwick,
Mathematics Institute & Department of Computer Science
charis@dcs.warwick.ac.uk

Abstract. In this paper we propose a deterministic algorithm for approximate counting the k-colourings of a graph G. We consider two categories of graphs. The first one is the sparse random graphs $G_{n,d/n}$, where each edge is chosen independently with probability d/n and d is fixed. The second one is a family we call *locally α-dense graphs* of bounded maximum degree Δ. A graph $G = (V, E)$ in this family has following property: For all $\{u, v\} \in E$ the number of vertices which are adjacent to v but not adjacent to u are at most $(1 - \alpha)\Delta$, where $\alpha \in [0, 1]$ is a parameter of the model.
Our algorithm computes in polynomial time a $(1 \pm n^{-\Omega(1)})$-approximation of the logarithm of the number of k-colourings of $G_{n,d/n}$ for $k \geq 2.1d$ with high probability. Also, for G a locally α-dense graph of *bounded* maximum degree Δ it computes in polynomial time a $(1 \pm (\log n)^{-\Omega(1)})$-approximation of the logarithm of the number of k-colourings, for $k > (2 - a)\Delta$. Restricting the treewidth of the neighbourhoods in G we get improved approximation.
Our algorithm is related to the algorithms of A. Bandyopadhyay et al. in SODA '06, and A. Montanari et al. in SODA '06. However, we use correlation decay properties for the *Gibbs distribution* in a completely different manner. I.e. given the graph $G = (V, E)$, we alter the graph structure in some specific region $\Lambda \subseteq V$ (by deleting edges between vertices of Λ) and then we show that the effect of this change on the marginals of Gibbs distribution, diminishes as we move away from Λ. Our approach is novel and suggests a new context for the study of deterministic counting algorithms.

1 Introduction

For a graph $G = (V, E)$ and a positive integer k, a proper k-colouring is an assignment $\sigma : V \to [k]$ (we use $[k]$ to denote $\{1, \ldots, k\}$), such that adjacent vertices receive different members of $[k]$, i.e. different colours. Here we focus on the problem of *counting* the k-colourings of G.

Typically, the set of colourings is exponentially large and exhaustive search is computationally prohibited. In [21] L. Valiant introduced the notion of $\#P$-

* Supported by EPSRC grant EP/G039070/1.

hardness and showed that exact counting of k-colourings is $\#P$-complete. Modulo a complexity theoretic conjecture the problems in $\#P$ do not admit a polynomial time algorithm. Research focuses on approximate counting algorithms.

Usually, a counting problem is reduced to computing certain marginal probabilities. Typically we estimate these marginals by using some *sampling* algorithm. The most powerful method for sampling is the Markov Chain Monte Carlo (MCMC). There the main technical challenge is to establish that the underlying Markov chain mixes in polynomial time (see [15]). The MCMC method gives *probabilistic* approximation guarantees.

Recently, new approaches were proposed for *deterministic* counting algorithms by [3] and [24]. This new approach links the correlation decay to *computing* efficiently marginals of Gibbs distributions. The correlation decay concept, also called spatial correlation decay or spatial mixing is related to notions such as uniqueness of Gibbs measures in infinite graphs (see [10]).

The two algorithms in [3, 24] suggest two different approaches for computing Gibbs marginals. In [3] the computation of marginals is done w.r.t. the graph-theoretic structure of the graph. In [24] the problem is reduced, first, to computing Gibbs marginals on some appropriately defined tree where some boundary conditions are imposed. These two different approaches require two different kinds of spatial correlation decay conditions.

The second approach gives better polynomial time approximations, however, it requires stronger correlation decay conditions. Attempts to establish such strong conditions were successful for "two-spin" cases, e.g. independent sets, matchings, e.t.c. (see [24, 4, 18]). For the "multi-spin" cases, such as colourings, things seem harder. The best algorithm of this category for counting k-colourings requires $k > 2.8\Delta$ and girth at least 4 (see [9]) [1].

In this paper we propose a deterministic algorithm for approximate counting the k-colourings of a graph G. The algorithm computes Gibbs marginals similarly to [3, 17], i.e. it fits to the first of the two approaches described above. However, our approach is novel, as we use spatial correlation decay in a completely different manner than that in [3, 17]. The spatial correlation conditions we require seem to be *weaker* than those in [3, 17].

In most of the cases considered, here, our algorithm outperforms the corresponding MCMC algorithms in terms of minimum colours required for polynomial-time execution.

Even though our approach is general and applies to any graph, here, we consider two, very interesting, cases: The first one is the *sparse random graph* $G_{n,d/n}$, where each edge is chosen independently with probability d/n and d is fixed. The second is a family of graphs that we call *locally α-dense graphs*.

Definition 1.1 (locally dense). *A graph $G = (V, E)$ is locally α-dense if the following holds: For all $\{u, v\} \in E$, v has at most $(1 - \alpha)\Delta$ neighbours which are not adjacent to u, where Δ is the maximum degree of G and $\alpha \in [0, 1]$ is a parameter of the model.*

[1] The best MCMC algorithm for sampling k-colourings of general graph, mixes in polynomial time for $k > 11/6\Delta$ (see [22]).

Note that the above family includes some $G = (V, E)$ having the following property: for all $v \in V$, the minimum degree of the subgraph induced by the neighbours of v is $\Theta(\Delta)$, where Δ is the maximum degree in G. However, the family of locally α-dense graphs is not restricted only to this kind of graphs, it is richer. In Section A.3 we show that this family allows graphs even with triangle free neighbourhoods. In this case the vertices of the triangle free neighbourhood should have sufficiently small degrees, i.e. $(1 - \alpha)\Delta$.

The *motivation* for studying the problem of counting k-colourings of graphs in these two categories is the following: As far as $G_{n,d/n}$ is concerned, the challenging task is the actual computation of Gibbs marginals for the k-colourings. Generally, the efficient computation of Gibbs marginals for the solutions of random constraint satisfaction problems has been central in Statistical Physics as well as in Computer Science, for the so called on average case analysis of the algorithms. Moreover, accurate computation of such marginals has been related to the empirical success of heuristics such as Survey Propagation and Belief Propagation in finding solution of instances of random constraint satisfaction problems (see [16]).

As far as locally dense graphs are regarded, we note that in the MCMC literature for sampling colourings, there is a large number of results showing that we can have polynomial mixing with $k < 11/6\Delta$ for special cases of graphs. In all these results, the common graph property required is that the neighbourhoods are fairly sparse, e.g. sufficiently large girth, planarity, e.t.c. (see [11, 13, 8]). Here we show that we can count colourings below the barrier of $11/6\Delta$, for bounded degree graphs, even when there exist very "dense" neighbourhoods.

1.1 Results

Let $Z(G, k)$ denote the number of k-colourings of the graph G. $Z(G, k)$ is also called *partition function*. Our algorithm computes an approximation for the log-partition function $\log Z(G, k)$.

Definition 1.2. *The value $\Psi(G, k)$ is defined to be ϵ-approximation of the log-partition function $\log Z(G, k)$ if*

$$(1 - \epsilon)\frac{\log Z(G, k)}{n} \leq \Psi(G, k) \leq (1 + \epsilon)\frac{\log Z(G, k)}{n}.$$

First we apply our algorithm on $G_{n,d/n}$ and we show the following theorem.

Theorem 1.1. *Taking sufficiently large d and $k > 2.1d$, then, with probability at least $1 - n^{-\epsilon_1}$ our counting algorithm* computes *a $n^{-\epsilon_2}$-approximation of $\log Z(G_{n,d/n}, k)$, in time $O(n^c)$, with ϵ_1, ϵ_2 and c positive real numbers which depend on k.*

Furthermore, we will show that the arithmetic value that is computed by the algorithm is with high probability[2] (w.h.p.) $n(\log\left(k(1 - 1/k)^{d/2}\right) \pm o(1))$. This

[2] With probability that tends to one as n tends to infinity.

result can, also, be derived by the second moment method, see the work of Achlioptas and Naor in [1]. Note that this result here is derived *algorithmically*, i.e. we provide a procedure which actually counts the number of colourings of $G_{n,d/n}$. Finally, we show that there is a set of graphs S_d with $Pr[G_{n,d/n} \in S_d] = 1 - n^{-0.1}$ such that for every $G \in S_d$ we can *verify in polynomial time* whether $\log Z(G,k)$ is equal to $n(\log\left(k(1-1/k)^{d/2}\right) \pm o(1))$ or not, when $k > 2.1d$.

For the locally dense graphs we get the following theorem.

Theorem 1.2. *For $G = (V,E)$ a locally α-dense graph of maximum degree Δ, a* fixed constant[3]*, and for $k > (2-\alpha)\Delta$ the counting algorithm computes a $(\log n)^{-\Omega(1)}$-approximation of $\log(Z(G,k))$ in polynomial time.*

For the non too restrictive case where $\alpha > 1/6$ the lower bound for k is less than $11/6\Delta$, i.e. less colours than what the best MCMC requires for polynomial-time sampling for this family of graphs. Restricting further the locally dense graph we get an even more accurate counting.

Corollary 1.1. *For $G = (V,E)$ a locally α-dense graph with the additional property that the $\Theta(\log n)$ neighbourhood of each $v \in V$ has bounded* treewidth*, the algorithm computes a $n^{-\Omega(1)}$-approximation of $\log(Z(G,k))$ in polynomial-time for $k > (2-\alpha)\Delta$.*

It is worth mentioning the least number of colours k that give polynomial mixing for the MCMC sampling for the two families of graphs we consider here. For $G_{n,d/n}$, k is a constant which depends on d with high probability (see [19]). For the second family k is the same as the general bound for MCMC sampling colourings of a graph with maximum degree Δ, i.e. $k > 11/6\Delta$ (see [22]).

We should note that the non MCMC algorithm for sampling k-colourings of $G_{n,d/n}$, with $k = k(d)$ constant[4] in [7], can be transformed so as to estimate the $\log Z(G_{n,d/n},k)$ in a same manner as in [3]. In this case, it should hold $k \geq d^3$ which is outperformed by this algorithm.

1.2 Contribution

As in [3, 17], the objective of our algorithm is to compute certain Gibbs marginals. The time for computing each of these marginals depends on the number of the vertices and the structure of the underlying graph G. Usually, this computation requires exponential time. Correlation decay is exploited to get polynomial time approximations for the marginals. In particular, correlation decay is used to restrict the necessary computations for each marginal to a small subgraph of G (i.e. instead of G we consider a small subgraph of G) where the computation is carried out in polynomial time. Crucially, the error in the estimation of the marginals depends on quantities which express spatial correlation decay.

In [3, 17] the correlation decay conditions that are used involve imposing boundary conditions (fixing the colouring of specific vertices) and showing that

[3] Δ is independent of the size of G.

[4] A very large constant that depends on d.

its effect on the Gibbs distributions decays as we move away from the boundary region. This approach is equivalent to establishing Dobrushin's Uniqueness Condition for infinite graphs (see [10]).

Here, we follow a completely different approach for the use correlation decay. The following example that illustrates the ideas that underlay our approach.

Example. Consider two graphs $G_a = (V, E_a)$ and $G_b = (V, E_b)$ and the region $\Psi \subset V$. Assume that $E_a \subset E_b$ while $E_b \backslash E_a$ includes only edges of G_b that are at graph distance at least t from Ψ. Let μ_a and μ_b be the Gibbs distributions of the k-colourings of G_a and G_b, correspondingly. For sufficiently large k, t we show that the marginal Gibbs distribution of the colourings of Ψ that is specified by μ_a and μ_b, correspondingly, are very close, in some sense.

Essentially, the correlation decay conditions we use here involve imposing a specific *graph structure* to some region by deleting edges and by showing that the effect of this change on the Gibbs distribution decays as we move away from the altered region.

The spatial correlation decay conditions we use seem to be *weaker* than those in [3, 17]. More specifically, we conjecture that our conditions are equivalent to what is known as *non-reconstructibility* conditions[5] for the colourings of G with a sufficiently large rate of decay (see [10]). We should remark that our conjecture would follow by showing that certain non-reconstructibility conditions are *monotone in the graph structure*. For further discussion see Section 3.1.

Theorem 3.1 and Theorem 3.2 provide the correlation decay conditions needed. However, in order to derive the results in Section 1.1 we need to provide appropriate bounds for the spatial correlation conditions. To achieve this we make a novel use of "Disagreement percolation" [5]. I.e. express the decay of correlation in terms of percolation-probabilities on graphs (see Section 3.2). This technique is general and simple, e.g. there is no need for restrictions on the graph structure. Furthermore, it allows to express the corresponding bounds in terms of the degree of each vertex, not the maximum degree [6]. Finally, in the most cases the correlation decay bounds will be *monotone in the graph structure*.

We should remark that one thing is the correlation decay conditions stated by Theorem 3.1 and Theorem 3.2 and an other thing is the bounds provided by the use of Disagreement Percolation.

1.3 Structure of the paper

The rest of the paper is organized as follows: In Section 2 we present some basic concepts and describe the counting problems w.r.t. these concepts. In Section 3 we present our counting algorithm and provide the theorems that relate its accuracy with certain spatial correlation decay. Also, we provide results which are used for bounding spatial correlation decay (in Section 3.2).

[5] Non-reconstructibility is equivalent to extrimality of Gibbs measure for infinite graphs, see e.g. [10].

[6] Property which will be exploited in the case of sparse $G_{n,p}$.

In the Appendix we provide in detail the proofs of the results presented in the main part of the paper. In particular, in Section A.1 we discuss the technical details for applying the counting algorithm on $G_{n,d/n}$. We prove Theorem 1.1 and other related results. We do the same for the locally α-dense graphs in Section A.2. In Section A.3 we present some remarks on locally dense graphs which provide a better view of this family of graphs.

In Section A.4 we prove the results that appear in Section 3.2 for bounding spatial correlation decay. In Section A.5 we give a detailed proof of the theorems and the lemmas that have not been proved to that point.

2 Basics and Problem Formulation

Let us introduce some notation, first. For the graph $G = (V, E)$ and the integer k let $\Omega(G, k)$ be the set of all proper k-colourings of G. We always assume that k is such that $Z(G, k) = |\Omega(G, k)| > 0$. Also, for $\Psi \subseteq V$ and $\sigma \in \Omega(G, k)$ we denote with $\Omega(G, k, \sigma_\Psi)$ the set of proper k-colourings of G which are consistent with σ on the colour assignments of the vertices in Ψ.

Our algorithm is studied in the context of finite spin-systems, a concept that originates in statistical physics. In particular, we use the finite colouring model.

The **Finite Colouring Model** with underlying graph $G = (V, E)$ that uses k colours is specified by a set of "*sites*", which correspond to the vertices of G, a set of "*spins*", i.e. the set $[k]$, and a symmetric function $U : [k] \times [k] \to \{0, 1\}$ such that for $i, j \in [k]$

$$U(i, j) = \begin{cases} 1 & \text{if } i \neq j \\ 0 & \text{otherwise.} \end{cases}$$

It holds that $|V| = n$ with n finite.

A *configuration* $\sigma \in [k]^V$ of the system assigns each vertex ("site")[7] $x \in V$ the colour ("spin value") $\sigma_x \in [k]$. The probability to find the system in configuration σ is determined by the *Gibbs distribution*, which is defined as

$$\mu(\sigma) = \frac{\prod_{\{x,y\} \in E} U(\sigma_x, \sigma_y)}{Z(G, k)} \tag{1}$$

where $Z(G, k)$ is the *partition function*. If and only if σ is a proper k-colouring of G, it holds $\prod_{\{x,y\} \in E} U(\sigma_x, \sigma_y) = 1$. This implies that the Gibbs distribution is the uniform distribution over $\Omega(G, k)$ [8].

Boundary condition corresponds to fixing the colour assignment of a specific "*boundary*" vertex set of G; the term "*free boundary*" implies that there is no boundary condition imposed.

[7] For convenience, when we refer to a site of a system we use the term vertex, while when we refer to a spin we use the term colour.

[8] The definition of Gibbs distribution is valid only for k so large that $|\Omega(G, k)| > 0$.

For convenience we denote with $PCS(G, k, \sigma_\Psi)$ the proper colouring model with underlying graph $G = (V, E)$ which uses k colours and there is a region $\Psi \subseteq V$ where we impose boundary conditions σ_Ψ, for $\sigma \in \Omega(G, k)$. The third parameter of PCS-notation is omitted, when we assume free boundary.

The probability for $PCS(G, k, \sigma_\Psi)$ [9] to be in configuration $\tau \in [k]^V$ is given by the corresponding Gibbs distribution μ^{σ_Ψ}, i.e.

$$\mu^{\sigma_\Psi}(\tau) \begin{cases} \frac{1}{Z(G,k,\sigma_\Psi)} & \text{if } \tau \in \Omega(G, k, \sigma_\psi) \\ 0 & \text{otherwise} \end{cases}$$

where $Z(G, k, \sigma_\Psi) = |\Omega(G, k, \sigma_\Psi)|$. If μ, μ^{σ_Ψ} are the Gibbs distributions of $PCS(G, k)$ and $PCS(G, k, \sigma_\Psi)$, correspondingly, then for $\tau \in [k]^V$ it holds that $\mu^{\sigma_\Psi}(\tau) = \mu(\tau|\sigma_\Psi)$.

Another concept we will need is that of the *sequence of subgraphs*.

Definition 2.1 (Sequence of subgraphs). *For the graph $G = (V, E)$, let $\mathcal{G}(G) = \{G_i = (V, E_i)\}_{i=0}^r$ denote a sequence of subgraphs of G which has the following properties:*

- *G_0 is some subgraph of G*
- *$E_i \subset E_{i+1}$ for $0 \leq i < r$ and $E_r = E$.*
- *the term G_{i+1} compared to G_i has an additional edge, the edge $\Psi_i = \{v_i, u_i\}$.*

When we refer to $\mathcal{G}(G)$ we specify the graph G_0 while we, usually, assume that there is some arbitrary rule which gives us some specific sequence of subgraphs.

For the graph G and some integer k such that $Z(G, k) > 0$, assume that we want to compute $Z(G, k)$. Consider a sequence of subgraphs $\mathcal{G}(G)$ such that G_0 is empty, e.g. $\mathcal{G}(G) = G_0, \ldots, G_r$. For each graph in $\mathcal{G}(G)$ consider the systems $PCS(G_i, k)$, for $0 \leq i \leq |E|$. For these systems, we define the following:

Definition 2.2. *W.r.t. to $\mathcal{G}(G)$ we define the sequence of systems $PCS(G_0, k)$,..., $PCS(G_r, k)$. Furthermore, let X_i denote the configuration of $PCS(G_i, k)$ and μ_i denote the Gibbs distribution specified by $PCS(G_i, k)$ for $0 \leq i \leq r$.*

For $\Lambda \subseteq V$, let $X_i(\Lambda) \in [k]^\Lambda$ be the random variable such that if $X_i = \sigma$, then $X_i(\Lambda) = \sigma_\Lambda$, for any $\sigma \in [k]^V$.

Lemma 2.1. *Considering the all above definitions, the following holds*

$$|Z(G, k)| = k^n \cdot \prod_{i=0}^{|E|-1} Pr[X_i(v_i) \neq X_i(u_i)] \tag{2}$$

where the vertices v_i and u_i are incident to Ψ_i (see Definition 2.1).

The proof of the above lemma appears in Section A.5.4.

The above lemma implies that we can have $Z(G, k)$ by computing the probability terms in (2). For the cases of graphs we are interested here these probability

[9] Where $\Psi \subseteq V$ and $\sigma \in \Omega(G, k)$

terms we have no polynomial-time method for their exact computation. Thus, we derive only polynomial-time approximations for this probabilities.

For the distributions ν_a, ν_b on $[k]^V$, let $||\nu_a - \nu_b||$ denote their *total variation distance*, i.e.

$$||\nu_a - \nu_b|| = \max_{\Omega' \subseteq [k]^V} |\nu_a(\Omega') - \nu_b(\Omega')|. \quad (3)$$

For $\Lambda \subseteq V$ let $||\nu_a - \nu_b||_\Lambda$ denote the total variation distance between the projections of ν_a and ν_b on $[k]^\Lambda$.

We close this section with some additional notation. For the graph $G = (V, E)$, the vertex $v \in V$ has degree Δ_v and the maximum degree of G is equal to Δ. Also, for $\Lambda \subseteq V$ and some integer $t > 0$ let $L(\Lambda, t) = \{v \in V | dist(\Lambda, v) = t\}$, where $dist(\Lambda, v)$ denotes the graph distance between Λ and v. Let, also, $B(\Lambda, t) = \cup_{t' \leq t} L(\Lambda, t')$ and $\bar{B}(\Lambda, t) = V \backslash B(\Lambda, t)$. Also, w.r.t. to the sequence of subgraph $G_0, G_1, \ldots, G_{r-1}$ let $L_i(\Lambda, t) = \{v \in V | dist(\Lambda, v) = t \text{ in } G_i\}$, similarly we define $B_i(\Lambda, t)$ and $\bar{B}_i(\Lambda, t)$.

3 Counting Schema

We are given a graph $G = (V, E)$ and some integer k such that $Z(G, k) > 0$. In this section we present our algorithm that estimates $\log Z(G, k)$.

Consider a sequence of subgraphs $\mathcal{G}(G) = G_0, \ldots, G_r$ with G_0 being empty. For a given integer t and every graph G_i in $\mathcal{G}(G)$ consider a, new, sequence of subgraphs $\mathcal{G}(G_i) = G_{i,0}, \ldots, G_{i,r_i}$ defined as follows: G_{i,r_i}, is the graph G_i while $G_{i,0}$ is derived from G_i be removing all the edges between the sets $L_i(\Psi_i, t)$ and $L_i(\Psi_i, t+1)$. We remind the reader that the intermediate terms of the sequences of subgaphs are constructed by some arbitrary rule. The value of t is a parameter of our algorithm.

W.r.t. the sequences $\mathcal{G}(G) = G_0, \ldots, G_r$ and $\mathcal{G}(G_i) = G_{i,0}, \ldots, G_{i,r_i}$, it should be clear to the reader what the symbols μ_i, $\mu_{i,j}$, X_i, $X_{i,j}$, Ψ_i and $\Psi_{i,j}$ for $0 \leq j \leq r_i$ and $0 \leq i \leq r$ stand for [10]. The counting algorithm follows:

Counting Schema

Input: G, k, t.
Set $\hat{Z}(G, k) = 1$.
For $0 \leq i \leq r - 1$ do

- Compute the exact value of $Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]$.
- Set $\hat{Z}(G, k) = \hat{Z}(G, k) \cdot Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]$.

End For.
Set $\hat{Z}(G, k) = \hat{Z}(G, k) \cdot k^n$.
Output: $\log\left(\hat{Z}(G, k)\right)/n$.

---

[10] Otherwise see Section 2

The above algorithm is based on the counting to marginal reduction stated in Lemma 2.1, while $Pr[X_i(v_i) \neq X_i(u_i)]$ is approximated by $Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]$. Note that the computation of $Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]$ is done w.r.t. the induced subgraph of G_i which includes only vertices that are within graph distance t from Ψ_i.

The computations for $Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]$ will involve graphs with a number of vertices of order either $\Theta(\log n)$ or a $O(n^{\Omega(1)})$. For the first case $Z(G_{i,0}, k)$ will be polynomially large and the computation of $Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]$ will be carried out by exhaustive enumeration. For the later case, $Z(G_{i,0}, k)$ may become exponentially large. There we require $G_{i,0}$ to have a special structure which allows a *dynamic programming algorithm* to carry out the computations in polynomial time. The actual choice of the algorithm is made ad-hoc.

Two natural questions arise for the counting algorithm. The first one is about its *accuracy*, i.e. how close $\frac{1}{n}\log \hat{Z}(G,k)$ and $\frac{1}{n}\log Z(G,k)$ are. The second one is about the *time complexity*. It should be straightforward that the execution time is dominated by the computations for $Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]$, for each i. When we consider a specific category of graphs we discuss in detail about the choice we make for the algorithm that computes the probability terms. In the remainder of this section we will focus on accuracy issues.

Theorem 3.1. *For the Counting Schema it holds that*

$$\frac{1}{n}|\log \hat{Z}(G,k) - \log Z(G,k)| \leq \frac{2}{n}\sum_{i=0}^{r-1} \frac{|Pr[X_i(v_i) \neq X_i(u_i)] - Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]|}{Pr[X_i(v_i) \neq X_i(u_i)]} \tag{4}$$

when each of the sumands in the r.h.s. is sufficiently small.

The proof of Theorem 3.1 appears in Section A.5.1.

So as to show that the estimation of $\log \hat{Z}(G,k)$ is accurate, typically, we do the following: We derive a constant lower bound for $Pr[X_i(v_i) \neq X_i(u_i)]$ in the denominator. Then, we show that $Pr[X_i(v_i) \neq X_i(u_i)]$ and $Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]$ are sufficiently close. There we use the following theorem.

Theorem 3.2. *For $0 \leq i \leq r-1$ it holds that*

$$|Pr[X_i(v_i) \neq X_i(u_i)] - Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]| \leq \sum_{j=1}^{r_i} C_{ij} \sum_{x \in \Psi_{i,j}} \max_{\sigma,\tau,\in\Omega(G_{ij},k)} ||\mu_{ij}(\cdot|\sigma_x) - \mu_{ij}(\cdot|\tau_x)||_{\Psi_i} \tag{5}$$

where $C_{ij} = C_{i,j}(G_{i,j}, k) = \max_{s,t\in[k]} \left\{(Pr[X_{i,j}(u_{i,j}) = s | X_{i,j}(v_{i,j}) = t])^{-1}\right\}$ and r_i is the number of terms in the sequence $\mathcal{G}(G_i)$.

The proof of Theorem 3.2 is given in Section A.5.2.

3.1 Remarks on the Spatial Conditions

It is interesting to discuss the implications of the spatial mixing conditions required by Theorem 3.1 and Theorem 3.2. Assuming that every C_{ij} is a sufficiently

small constant, which will be the case here, (5) suggests that

$$\frac{1}{n}|\log \hat{Z}(G,k) - \log Z(G,k)| \leq f(G,t) \cdot \max_{i,j,x,\sigma,\tau} ||\mu_{ij}(\cdot|\sigma_x) - \mu_{ij}(\cdot|\tau_x)||_{\Psi_i} \quad (6)$$

where $f(G,t)$ is a quantity that grows linearly with the number of summads in (4) and (5). It is direct that a sufficient condition for the counting schema to be efficient and accurate is that for every $0 \leq i \leq r$ and $0 \leq j \leq r_i$ to have

$$\max_{x \in V \sigma,\tau \in \Omega(G_{ij},k)} ||\mu_{ij}(\cdot|\sigma_x) - \mu_{ij}(\cdot|\tau_x)||_{L_{ij}(\{x\},t)} \leq C_0 \cdot \exp\{-a_0 \cdot t\} \quad (7)$$

for appropriate fixed real numbers $C_0 > 0$ and $a_0 > 0$. Clearly, the above inequality makes sense when $f(G,t)$ in (6) grows with t slower than the inverse of the r.h.s. in (7), i.e. the quantity a_0 in (7) is sufficiently large. We also have the following.

Lemma 3.1. *For any graph $G = (V,E)$ and k, let μ be the Gibbs distribution of $PCS(G,k)$. For every $x \in V$ and $\Lambda \subseteq V$ it holds*

$$\sum_{A \in [k]^\Lambda} \mu(A) \cdot ||\mu(\cdot|A) - \mu(\cdot)||_x \leq \max_{\sigma,\tau \in \Omega(G,k)} ||\mu(\cdot|\sigma_x) - \mu(\cdot|\tau_x)||_\Lambda \leq 2k \cdot \sum_{A \in [k]^\Lambda} \mu(A) \cdot ||\mu(\cdot|A) - \mu(\cdot)||_x.$$

For a proof Lemma 3.1 see in Section A.5.5.

By Lemma 3.1 the condition in (7) is equivalent to the following: For $0 \leq i \leq r$ and $0 \leq i \leq r_i$

$$\max_{x \in V \sigma,\tau \in \Omega(G_{ij},k)} \sum_{A \in [k]^{L_{ij}(\{x\},t)}} \mu_{ij}(A) \cdot ||\mu_{ij}(\cdot|A) - \mu_{ij}(\cdot)||_x \leq C \exp\{-a \cdot t\} \quad (8)$$

for appropriate fixed $C, a > 0$. The above is known as *non-reconstructibility* condition for the k-colourings of $G_{i,j}$ (with exponential decay). This condition implies that a "typical" colouring of $L_{i,j}(\{x\},t)$ should have small impact to the Gibbs marginal on x.

Under the assumption that the *non-reconstructibility* condition is *monotone* in the graph structure, the condition in (8) is equivalent to the condition

$$\max_{x \in V} \sum_{A \in [k]^{L(\{x\},t)}} \mu(A) \cdot ||\mu(\cdot|A) - \mu(\cdot)||_x \leq C \exp\{-a \cdot t\} \quad (9)$$

where μ is the Gibbs distribution specified by $PCS(G,k)$. Note that the above condition is weaker than the one required in [3, 7, 17] i.e.

$$\max_{\sigma \in \Omega(G,k)} ||\mu(\cdot|\sigma_{L(\{x\},t)}) - \mu(\cdot)||_x \leq C' \exp\{-a' \cdot t\} \quad (10)$$

for fixed real numbers $C' > 0$ and $a' > 0$. Note that the above condition requires that the effect of any "arbitrary" colouring of $L(\{x\},t)$ should have small impact on the Gibbs marginal on x.

To our knowledge establishing such a monotonicity property for the non-reconstructibility condition is an *open problem*.

3.2 Bounds for Spatial Correlation decay

In this section we provide two theorems that can be used to derive upper bounds for the quantities $||\mu_{ij}(\cdot|\sigma_x) - \mu_{ij}(\cdot|\tau_x)||_{\Psi_i}$in (5). The derivation of the bounds are of independent interest than the discussion in the Section 3.1. The technique that is used by the two theorems is the well known "*disagreement percolation*" coupling construction, [5]. Here, we present only the statements of the theorems while their proofs are presented in Section A.4.

Note that the disagreement percolation allows to express the quantities in the r.h.s of (5) in percolation-like probabilities, which in turn can be expressed in terms of the number of colours k and the degree Δ_v of each $v \in V$ in G_{ij}. *To avoid the notation with the indices i, j, we consider a general graph $G = (V, E)$, instead.*

Let $\mathcal{P}_{\boldsymbol{q}}$, for $\boldsymbol{q} \in [0,1]^V$, denote a product measure under which the vertex $v \in V$ is *disagreeing* with probability q_v and *non-disagreeing* with probability $1 - q_v$. A *path of disagreement* in G is any simple path which has all its vertices disagreeing.

Theorem 3.3. *For the graph $G = (V, E)$ let μ be the Gibbs distribution specified by $PCS(G, k)$. W.r.t. G, $x \in V$ and k let $\mathcal{P}_{\boldsymbol{q}}$ denote the product measure such that for $v \in V\backslash\{x\}$ of degree Δ_v, we have $q_v = \frac{1}{k-\Delta_v}$ while $q_x = 1$. It holds that*

$$\max_{\sigma,\eta\in\Omega(G,k)} ||\mu(\cdot|\sigma_x) - \mu(\cdot|\eta_x)||_{\Lambda} \leq \mathcal{P}_{\boldsymbol{q}}[\ \exists \text{ path of disagreement connecting } \{x\} \text{ and } \Lambda \text{ in } G]. \tag{11}$$

where $\Lambda \subseteq V$.

The proof of Theorem 3.3 is given in Section A.4.

Let $\mathcal{BP}(x, \Lambda)$ be the set of all paths connecting x with Λ such that the vertices of each path form an induced subgraph of G which is bipartite. The analysis in the proof of Theorem 3.3 suggests that we can restrict the "disagreement" to the paths in $\mathcal{BP}(x, \Lambda)$. For each $M \in \mathcal{BP}(x, \Lambda)$ let

$$\mathcal{I}_M = \begin{cases} 1 & \text{if } M \text{ is a path of disagreement} \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 3.4. *For the graph $G = (V, E)$ let μ be the Gibbs distribution that is specified by the system $PCS(G, k)$. W.r.t. G and any $x \in V$, consider the product measure $\mathcal{P}_{\boldsymbol{q}}$, with $\boldsymbol{q} \in [0,1]^V$, such that for $v \in V\backslash\{x\}$ of degree Δ_v, we have $q_v = \frac{1}{k-\Delta_v}$ while $q_x = 1$. It holds that*

$$||\mu(\cdot|\sigma_x) - \mu(\cdot|\eta_x)||_{\Lambda} \leq \sum_{M\in\mathcal{BP}(x,\Lambda)} E_{\boldsymbol{q}}[\mathcal{I}_M]. \tag{12}$$

where $\Lambda \subseteq V$ and $E_{\boldsymbol{q}}[\cdot]$ is the expectation taken w.r.t. to probability measure $\mathcal{P}_{\boldsymbol{q}}$.

The proof of Theorem 3.4 is, also, presented in Section A.4.

Acknowledgement. The author would like to thank Amin Coja-Oghlan for the numerous fruitful discussions and for the encouragement to study this problem.

References

1. D. Achlioptas, and A. Naor. *The two possible values of the chromatic number of a random graph.* In proc. of STOC' 04, pp 587-593, 2004.
2. D. Aldous. *Random walks of finite groups and rapidly mixing Markov chains.* In: Séminaire de Probabilités XVII 1981/82, Springer-Verlag, Berlin. pp. 243-297.
3. A. Bandyopadhyay and D. Gamarnik. *Counting without sampling: new algorithms for enumeration problems using statistical physics.* In proc. of SODA '06 pp 890-899, 2006.
4. M. Bayati, D. Gamarnik, D. Katz, C. Nair, and P. Tetali. *Simple deterministic approximation algorithms for counting matchings.* In Proc. of STOC '07. pp 122-127, 2007.
5. J. van den Berg and C. Maes. *Disagreement percolation in the study of Markov fields.* Annals of Probability 22 pp. 794-763 1994.
6. M. Dyer, A. Flaxman, A. M. Frieze and E. Vigoda. *Random colouring sparse random graphs with fewer colours than the maximum degree.* Random Struct. and Algorithms 29, pp 450-465, 2006.
7. C. Efthymiou and P. G. Spirakis. *Random sampling of colourings of sparse random graphs with a constant number of colours.* In Theoretical Computer Science 407, pp. 134-154, 2008.
8. A. M. Frieze and J. Vera. *On randomly colouring locally sparse graphs.* In J. Discrete Mathematics & Theoretical Computer Science 8 (1), pp 121-128, 2006
9. D. Gamarnik and D. Katz. *Correlation decay and deterministic FPTAS for counting list-colorings of a graph.* In Proc. of SODA 2007.
10. H. O. Georgii. *Gibbs measures and phase transitions*, de Gruyter Studies in Mathematics 9, Walter de Gruyter & Co., Berlin, 1988.
11. L. A. Goldberg R. A. Martin and M. Paterson. *Strong Spatial Mixing with Fewer Colors for Lattice Graphs* In SIAM Journal of Computing 35 (2) pp486-517, 2005.
12. O. Häggström. *A note on disagreement percolation.* In Random Structures and Algorithms 18 pp. 267-278 2001.
13. T. P. Hayes, J. C. Vera and E. Vigoda. *Randomly coloring planar graphs with fewer colors than the maximum degree* In Proc. of STOC '07 pp. 450-458, 2007.
14. S. Janson, T. Luczak and A. Ruciński. *Random graphs* Wiley and Sons, Inc. 2000.
15. M. Jerrum and A. Sinclair. *The Makov chain Monte Carlo method: an approach to approximate counting and integration.* In Approximation Algorithms for NP-hard problems (Dorit, Hochbaum ed.) PWS 1996.
16. F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjianc, L. Zdeborova. *Gibbs states and the set of solutions of random constraint satisfaction problems.* In Proc. National Academy of Sciences 104 (2007) 10318-10323.
17. A. Montanari and D. Shah. *Counting good truth assignments of random k-SAT formulae.* In proc. of SODA 07, pp 1255-1264, 2007.
18. E. Mossel and A. Sly. *Rapid Mixing of Gibbs Sampling on Graphs that are Sparse on Average.* In proc. of SODA'08, pp 238-247, 2008.
19. E. Mossel and A. Sly. *Gibbs Rapidly Samples Colorings of* $G_{n,d/n}$. Preprint http://arxiv.org/abs/0707.3241 (To appear in journal PTRF.)
20. B. A. Reed. *Finding approximate separators and computing treewidth quickly.* In Proc. STOC '92. pp 221-228 2008.
21. L. Valiant. *The complexity of enumeration and reliability problems.* SIAM Journal of Computing 8, pp 410-421, 1979.

22. E. Vigoda. *Improved bounds for sampling colorings.* In J. Math. Phys. (41), pp 1555-1569, (2000).
23. M. Wainwright and M. Jordan. *Graphical Models, exponential families and variational inference*. Technical report 649, Dept. of Statistics, University of Berkeley.
24. D. Weitz. *Counting independent sets up to the tree threshold.* In Proc. of the 38th annual Symposium of Theory of Computing (STOC 06) 2006.

Appendix

A.1 Application - $G_{n,d/n}$ case

In this section we show Theorem 1.1 and the results related to $G_{n,d/n}$. For technical reasons, which we discuss later, require the following sequence of subgraphs.

Sequence of subgraphs $\mathcal{G}(G_{n,d/n})$: Let r be the greatest index in $\mathcal{G}(G_{n,d/n})$, e.g. $\mathcal{G}(G_{n,d/n}) = G_0, \ldots, G_r$. The term G_0 is empty. Let R be the set of all edges in $G_{n,d/n}$ that do not belong to a cycle of length more than $\frac{\log n}{10 \log(e^2 d/2)}$ but they are adjacent to a vertex that belongs to such a cycle. There is an index i_0 such that for $i \geq i_0$, G_i differs from G_{i-1} in some edge from R, in $\mathcal{G}(G_{n,d/n})$.

For $0 \leq i \leq r$ consider that the sequence of subgraphs $\mathcal{G}(G_i)$ defined as follows: $G_{i,0}$ is derived by G_i be deleting all the edges that connect the sets of vertices $L_i(\Psi_i, a \log n)$ and $L_i(\Psi_i, a \log n + 1)$ where $0 < a < \frac{1}{2 \log d}$.

Due to the fact that k is smaller than the maximum degree of $G_{n,d/n}$ [11] there can be cases that $(C_{i,j})^{-1}$ and $Pr[X_i(v_i) \neq X_i(u_i)]$ are very small. This is undesired since it may have bad effect on the accuracy of the counting.

The proof of Theorem 1.1 implies that this may occur when the vertices which are involved, i.e. v_i, u_i, v_{ij}, u_{ij}, have large degrees and belong to small cycles. To avoid such cases, for every edge $\{v, u\}$ that is adjacent to a sufficiently small cycle the estimation of $Pr[X(v) \neq X(u)]$ is set $(1 - 1/k)$. Working in this way, all the "small" cycles that appear $G_{n,d/n}$ become isolated in G_i, for $i < i_0$. This forbids any case of having exceedingly small either $(C_{i,j})^{-1}$ or $Pr[X_i(v_i) \neq X_i(u_i)]$, for $i < i_0$. Furthermore, the corresponding error introduced in the estimation of $\log(Z(G_n, d/n, k))$ will be negligible.

Note that we set the parameter t of the Counting Schema equal to $a \log n$, for $0 < a < \frac{1}{2 \log d}$. For such t, the component in $G_{i,0}$ which contains $\{v_i, u_i\}$ is w.h.p. a tree with at most $\Theta(\log n)$ extra edges, where $0 \leq i < i_0$. This allows the schema to run in polynomial-time.

Computing Probabilities. The probability term $Pr[X_{i,0}(v_i) \neq X_{i,0}(v_i)]$, for $0 \leq i < i_0$, can be computed by using Dynamic Programming.

More specifically, note that using DP we can compute exactly the number of list colourings of a tree T. In the list colouring problem every vertex $v \in T$ has a set $List(v)$ of valid colours, where $List(v) \subseteq [k]$ and v only receives a colour in $List(v)$. For a tree on l vertices, using dynamic programming we can compute exactly the number of list colourings in time lk.

The connected components in $G_{i,0}$ that contain $\{v_i, u_i\}$ is a tree with at most $\Theta(\log n)$ extra edges, for $0 \leq i < i_0$, w.h.p.. For such component we can consider all $\leq k^{\Theta(\log n)}$ colourings of the endpoints of the extra edges and for each of these colourings recurse on the remaining tree. Since in our case k is a constant, $k^{\Theta(\log n)}$ is a polynomial of n. Thus, w.h.p. the numer of list colourings of the connected component, in $G_{i,0}$, that contains $\{v_i, u_i\}$ can be counted in polynomial time for every i. It is direct that this is sufficient for computing $Pr[X_{i,0}(v_i) \neq X_{i_0}(u_i)]$ [12].

[11] The maximum degree in $G_{n,d/n}$ is $\Theta\left(\frac{\log n}{\log \log n}\right)$ w.h.p. (see [14])

[12] A similar DP approach is also used in [6] and [7].

In what follows, we provide the pseudocode of the counting schema for $G_{n,d/n}$. Note that, the input parameter ϵ controls the quality of the approximation of $\log(Z(G_{n,d/n}, k))$.

Counting Schema $G_{n,d/n}$

Input: $G_{n,d/n}$, k, ϵ
If $k \leq 2.1d$ compute $\log(Z(G_{n,d/n}, k))$ by exhaustive enumeration.
If $\epsilon < n^{-a \log(k/2d)}$ compute $\log(Z(G_{n,d/n}, k))$ by exhaustive enumeration.
Compute the set of edges R.
If $|R| > n^{0.3}$ make an exhaustive enumeration of colourings.
Compute the sequence of subgraphs $\mathcal{G}(G_{n,d/n})$.
Set $\hat{Z}(G, k) = \left(1 - \frac{1}{k}\right)^{|R|}$.
For $0 < i < r - |R|$ do

- Compute the exact value of $Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]$.
- Set $\hat{Z}(G, k) = \hat{Z}(G, k) \cdot Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]$.

End for.
Set $\hat{Z}(G, k) = \hat{Z}(G, k) \cdot k^n$.
Output: $\log\left(\hat{Z}(G, k)\right)/n$.

Theorem 1.1 follows as a corollary of the following two theorems.

Theorem A.1.1. *For sufficiently large d, $k > 2.1d$ the counting schema computes a $n^{-\epsilon_2}$-approximation of $\log Z(G_{n,d/n}, k)$, with probability at least $1 - n^{-\epsilon_1}$, where $\epsilon_1 > 0$ and $\epsilon_2 > 0$ depend on k.*

The proof of Theorem A.1.1 appears in Section A.1.1 and makes a heavy use of Theorem 3.3.

Theorem A.1.2. *There are real constants $c > 0$ and $\epsilon > 0$ such that the time complexity for the counting schema to compute the log-partition $G_{n,d/n}$ is $O(n^c)$, with probability at least $1 - n^{-\epsilon}$.*

Proof: Theorem follows directly from the paragraph, "Computing Probabilities", above. ◇

Furthermore, the proof of Theorem A.1.1 implies the following theorem too.

Theorem A.1.3. *For sufficiently large d and $k > 2.1d$, w.h.p. it holds that*

$$\left| \frac{\log Z(G_{n,d/n}, k)}{n} - \left(\log k + \frac{d}{2} \cdot \log\left(1 - \frac{1}{k}\right) \right) \right| \leq o(1). \tag{13}$$

If d is relatively small, then the above relation holds w.h.p. when $k = \Theta(d)$.

The proof of Theorem A.1.3 appears in Section A.1.2.

Theorem A.1.4. *For fixed d and integer $k > 2.1d$, there exist a set of graphs S_d with $Pr[G_{n,d/n} \in S_d] = 1 - n^{-0.1}$ and a function $h(n, d, k) = n^{-\Omega(1)}$ such that for every $G \in S_d$ it can be verified in polynomial-time whether the property*

$$\left| \frac{\log Z(G, k)}{n} - \left(\log k + \frac{d}{2} \cdot \log\left(1 - \frac{1}{k}\right) \right) \right| \leq h(n, k, d). \tag{14}$$

holds or not.

The proof of Theorem A.1.4 appears in Section A.1.3

A.1.1 Proof of Theorem A.1.1

First we present a series of lemmas and corollaries that will be used for proof of Theorem A.1.1. In essence in these preliminary results we derive appropriate bounds for the quantities $||\mu_{ij}(\cdot|\sigma_x) - \mu_{ij}(\cdot|\tau_x)||_{\Psi_i}$, C_{ij} (defined in Theorem 3.2) and $Pr[X_i(v_i) \neq X_i(u_i)]$, for all i, j, by making extensive use of Theorem 3.3.

W.r.t. to the graph $G_{n,d/n}$ and the number of colours k, consider the product measure $\mathcal{P}_{\boldsymbol{q}}$, as defined in the statement of Theorem 3.3.

Definition A.1.1. *Let $q(s, d/n)$ be the following quantity*

$$q(s,d/n) \leq \sum_{i=0}^{s-1} \frac{1}{s-i}\binom{n}{i}\left(\frac{d}{n}\right)^i\left(1-\frac{d}{n}\right)^{n-i} + \sum_{i=s}^{n}\binom{n}{i}\left(\frac{d}{n}\right)^i\left(1-\frac{d}{n}\right)^{n-i} \tag{15}$$

Clearly, $q(s, d/n)$ is an upper bound for the marginal probability of the event that a vertex in $G_{n,d/n}$ is "disagreeing" under the measure $\mathcal{P}_{\boldsymbol{q}}$ when s colours are used. For convenience, we will use $q(s, d/n)$ with s not an integer, then we assume that we have $q(\lceil s \rceil, d/n)$.

Lemma A.1.1. *Consider the product measure $\mathcal{P}_{\boldsymbol{q}}$, w.r.t. the graph $G_{n,d/n}$ and number of colours k. Let π be a permutation of l vertices of $G_{n,d/n}$, for $0 \leq l \leq \Theta(n^{1/2})$. Conditioning that the maximum degree of $G_{n,d/n}$ is at most $\log^2 n$, for sufficiently large n it holds that*

$$\mathcal{P}_{\boldsymbol{q}}[\pi \textit{ is a path of disagreement}] \leq q(k,d/n)\cdot\left(\frac{d}{n}\right)^l\cdot\left(q(k-d/100,d/n)+n^{-0.499}\right)^l(1+o(1)).$$

Proof: Call π the path that corresponds to the permutation π, e.g. $\pi = (x_1, \ldots x_l)$. Let Γ be such that

$$\Gamma = \begin{cases} 1 & \text{if } \pi \text{ is a path of disagreement} \\ 0 & \text{otherwise.} \end{cases}$$

Note that the probability for $G_{n,d/n}$ to have a vertex of degree greater than $\log^2 n$ is less than $n^{-\omega(n)}$. It holds that

$$E_{\boldsymbol{q}}[\Gamma] \leq \left(\frac{d}{n} + n^{-\omega(n)}\right)^l \cdot E_{\boldsymbol{q}}[\Gamma|\text{the path } \pi \text{ appears in } G_{n,d/n}].$$

By the fact that $(1+n^{-\omega(n)})^l \leq \exp\left\{n^{-\omega(n)}n^{1/2}\right\} = 1 + o(1)$, we get

$$E_{\boldsymbol{q}}[\Gamma] \leq (1+o(1)) \cdot \left(\frac{d}{n}\right)^l \cdot E_{\boldsymbol{q}}[\Gamma|\text{the path } \pi \text{ appears in } G_{n,d/n}].$$

It suffice to show that for $0 \leq l \leq n^{1/2}$ and sufficiently large n, that

$$E_{\boldsymbol{q}}[\Gamma|\text{the } \pi \text{ appears in } G_{n,d/n}] \leq q(k,d/n)\cdot\left(q(k-\lceil d/100\rceil, d)+n^{-0.499}\right)^l. \tag{16}$$

We show (16) by induction on l. Clearly for $l = 0$, (16) is true, Assuming that (16) holds for $l = l_0$, we will show that it also holds for $l = l_0 + 1$, as well.

We assume that the path $(x_1, \ldots, x_{l_0+1})$ appears in $G_{n,d/n}$. For any vertex u, let D_u, denote the event that the vertex u is disagreeing. Given that all vertices in $\{x_1, \ldots, x_{l_0}\}$ are disagreeing let $\hat{\Delta}_{x_i}$ be the number of vertices in $V \backslash \{x_1, \ldots, x_{l_0}\}$ that are adjacent to x_i, for $1 \leq i \leq l_0$.

Note that if $\hat{\Delta}_{x_i} = t$, then all the possible subsets of $V \backslash \{x_1, \ldots, x_{l_0}\}$ of cardinality t are equiprobable to be adjacent to x_i. This implies that the probability for x_{l_0+1} to be adjacent to x_i is $\frac{E[\hat{\Delta}_{x_i}]}{n-l_0}$.

Let $\tilde{\Delta}_{l_0+1}$ be the number of edges that are adjacent to x_{l_0+1} and some vertex in $\{x_1, \ldots, x_{l_0-1}\}$. By the linearity of expectation we have

$$E[\tilde{\Delta}_{l_0+1}] \leq \frac{l_0}{n-l_0} E[\hat{\Delta}_{x_i}] \leq C_0 n^{-1/2} \log^2 n \tag{17}$$

for a sufficiently large real C_0. The last inequality follows from the assumption that $0 < l_0 < n^{1/2}$, the fact that the maximum degree is at most $\log^2 n$ and the fact that $E[\hat{\Delta}_{x_i}]$ is less than the maximum degree, for $i = 0, \ldots, l_0$.

We make the simplifying assumption that if the vertex x_{l_0+1} is adjacent to more than $d/100$ vertices in $\{x_1, \ldots, x_{l_0-1}\}$, then it is disagreeing, regardless of the number of adjacent vertices outside the path. By (17) and the Markov inequality, for sufficiently large n we get that

$$Pr[\tilde{\Delta}_{l_0+1} > d/100] \leq \frac{C_0 \cdot d}{100} n^{-1/2} \log^2 n \leq n^{-0.499}.$$

Remark. Given that $\{x_1, \ldots, x_{l_0}\}$ is a path of disagreement, the number of edges between x_{l_0+1} and the set of vertices $V \backslash \{x_1, \ldots, x_{l_0}\}$ is dominated by $\mathcal{B}(n, d/n)$[13].

We denote with E the event that "$(x_1, \ldots, x_{l_0})$ *is a path of disagreement,* $\tilde{\Delta}_{x_{l_0+1}} < d/100$*, the edge* $\{x_{l_0}, x_{l_0+1}\}$ *appears in* $G_{n,d/n}$". The following inequality is straightforward.

$$\begin{aligned} Pr[D_{x_{l_0+1}} | E] \leq & \sum_{j=0}^{k-\tilde{\Delta}_{x_{l_0+1}}+1} \frac{1}{k-j-\tilde{\Delta}_{x_{l_0+1}}} \binom{n}{j} (d/n)^j (1-d/n)^{n-j} + \\ & + \sum_{j=k-\tilde{\Delta}_{x_{l_0+1}}}^{n} \binom{n}{j} (d/n)^j (1-d/n)^{n-j} \\ & \leq q(k - d/100, d/n). \end{aligned}$$

The lemma follows. ◇

Lemma A.1.2. *For sufficiently large* d *and* $k \geq 2.1d$*, consider the graph* $G_{n,d/n}$ *and let* μ *be the Gibbs distribution specified by* $PCS(G_{n,d/n}, k)$*. For every* $x \in V$ *and* $\Psi \subseteq L(x, l)$*, for* $0 \leq l \leq \Theta(n^{1/2})$*, there exists a real* $\gamma > 1$ *such that*

$$E\left[\max_{\sigma,\eta \in \Omega(G_{n,d/n},k)} ||\mu(\cdot|\sigma_x) - \mu(\cdot|\eta_x)||_\Psi\right] \leq C \cdot \frac{|\Psi|}{n} \gamma^{-l} \tag{18}$$

where $|\Psi|$ *is the number of vertices in* Ψ *and* C *is a sufficiently large real constant. Note that the expectation is over the graphs.*

[13] Actually it is dominated by $\mathcal{B}(n - l_0, d/n)$ which in turn is dominated by $\mathcal{B}(n, d/n)$

Proof: Consider an enumeration of all the permutations of l vertices in $G_{n,d/n}$ with first the vertex x and last some vertex of Ψ. Let $\pi_0(l), \pi_1(l), \ldots$ be the permutations in the order they appear in the enumeration. We will say that the l-path i appears in $G_{n,d/n}$ if and only if $G_{n,d/n}$ contains a path whose vertices appear in the order specified by $\pi_i(l)$.

W.r.t. the graph $G_{n,d/n}$, consider the product measure $\mathcal{P}_{\boldsymbol{q}}$ as it is defined in the statement of Theorem 3.3. Let $\Gamma_i(l)$ be the random variable such that

$$\Gamma_i(l) = \begin{cases} 1 & \text{the path that corresponds to } \pi_i(l) \text{ is a path of disagreement} \\ 0 & \text{otherwise.} \end{cases}$$

Let, also, $\Gamma(l) = \sum_i \Gamma_i(l)$. By Theorem 3.3 we have that for a given instance of $G_{n,d/n}$ it holds

$$\max_{\sigma,\eta\in\Omega(G_{n,d/n},k)} ||\mu(\cdot|\sigma_x) - \mu(\cdot|\eta_x)||_\Psi \leq \mathcal{P}_{\boldsymbol{q}}\left[\cup_{l>0} \cup_i \Gamma_i(l)\right] \leq \sum_{i,l} \mathcal{P}_{\boldsymbol{q}}\left[\Gamma_i(l)\right]$$

Taking the average over both the graph instances and the product measure $\mathcal{P}_{\boldsymbol{q}}$ we get that

$$E\left[\max_{\sigma,\eta\in\Omega(G_{n,d/n},k)} ||\mu(\cdot|\sigma_x) - \mu(\cdot|\eta_x)||_\Psi\right] \leq \sum_{i,l} E[\Gamma_i(l)]. \tag{19}$$

Lemma will follow by bounding appropriately the r.h.s of (19).

For the moment assume that the maximum degree in $G_{n,d/n}$ is at most $\log^2 n$. Later in this proof we remove this assumption. We should mention that the probability for the event that the maximum degree in $G_{n,d/n}$ is at most $\log^2 n$ is at least $1 - n^{-\omega(n)}$.

Using Lemma A.1.1 and the fact that there are less than $|\Psi| \cdot n^{l-1}$ paths of length l, between x and Ψ, we get

$$E_{\boldsymbol{q}}[\Gamma(l)] = \sum_i E_{\boldsymbol{q}}[\Gamma_i(l)] \leq \frac{|\Psi|}{n} d^l \cdot \left(q(k - d/100, d/n) + n^{-0.499}\right)^l.$$

Setting $s = k - d/100$, where $k > 2.1 \cdot d$ we get

$$\begin{aligned} q(s, d/n) &\leq \sum_{j=0}^{s-1} \frac{1}{s-j}\binom{n}{j}(d/n)^j(1-d/n)^{n-j} + \sum_{j=s}^{n}\binom{n}{j}(d/n)^j(1-d/n)^{n-j} \\ &\leq \frac{2}{s}\sum_{j=0}^{s/2}\binom{n}{j}(d/n)^j(1-d/n)^{n-j} + \sum_{j=s/2+1}^{n}\binom{n}{j}(d/n)^j(1-d/n)^{n-j} \\ &\leq \frac{2}{s} + z(s/2). \end{aligned}$$

The quantity $z(s/2)$ is equal to the probability for a random variable distributed as in $\mathcal{B}(n, d/n)$ to be greater than $s/2$. Using Chernoff bounds and for sufficiently large d, it holds $z(s/2) < d^{-2}$. Thus, we get that $q(s, d/n) + n^{-0.49} \leq \frac{1}{1.04 \cdot d}$ and

$$E_{\boldsymbol{q}}[\Gamma(l)] \leq \frac{|\Psi|}{n}(1.04)^{-l}.$$

With similar arguments we get that the expected number of paths of disagreement of length l that start from x and get to *any* vertex in $G_{n,d/n}$ is $(1.04)^{-l}$.

Let E_s = *"there is no path of disagreement that starts from x and has length $\frac{\hat{C}\log n}{\log(1.04)}$"*, where $\hat{C} > 0$ is a sufficiently large constant. By the Markov Inequality we upper bound the probability of the complementary of E_s, i.e. $\mathcal{P}_{\boldsymbol{q}}[\bar{E}_s] \leq n^{-\hat{C}}$. Thus, $\mathcal{P}_{\boldsymbol{q}}[E_s] \geq 1 - n^{-\hat{C}}$. Also, it is easy to see that $E_{\boldsymbol{q}}[\Gamma(l)|E_s] \leq 2E_{\boldsymbol{q}}[\Gamma(l)]$.

Let Γ be the number of paths that start from x, have length at most $\frac{\hat{C}\log n}{\log(1.04)}$ and connect the vertex x with Ψ. Then

$$E_{\boldsymbol{q}}[\Gamma|E_s] \leq 2\frac{|\Psi|}{n} \sum_{l=\epsilon\log n}^{\hat{C}\log n/\log(1.04)} E_{\boldsymbol{q}}[\Gamma(l)] \leq C_1 \frac{|\Psi|}{n} (1.04)^{-\epsilon\log n}$$

where C_1 is a sufficiently large positive constant. Call D the event that the maximum degree in $G_{n,d/n}$ is less than $\log^2 n$. Let

$$S = \max_{\sigma,\eta\in\Omega(G,k)} ||\mu(\cdot|\sigma_x) - \mu(\cdot|\eta_x)||_\Psi.$$

It holds that

$$E[S] = E[S \mid D, E_s]\, Pr[D, E_s] + E\left[S \mid \bar{D} \cup \bar{E}_s\right] Pr[\bar{D} \cup \bar{E}_s].$$

Clearly

$$E[S \mid D, E_s] \leq E[\Gamma|D, E_s] \leq C_1 \frac{|\Psi|}{n} (1.04)^{-\epsilon\log n}.$$

Also, it holds that $E\left[S \mid \bar{D} \cup \bar{E}_s\right] \leq 1$. Thus we get that

$$E[S] \leq C_1 \frac{|\Psi|}{n} (1.04)^{-\epsilon\log n} + n^{-\omega(n)} + n^{-\hat{C}}.$$

Lemma follows by taking sufficiently large $\hat{C}$. ◇

Corollary A.1.1. *The bound for the expectation in (18) holds even if we remove an arbitrary set of edges of $G_{n,d/n}$.*

Proof: This holds since we have used Theorem 3.3 whose bounds are increasing in the degrees of the vertices. ◇

Lemma A.1.3. *With probability at least $1 - n^{-0.19}$ the number of edges in R are at most $n^{0.3}$, while the shortest cycle in $G_{n,d/n}$ which contain each of them is of length greater than $\frac{\log n}{10\log(e^2 d/2)}$.*

Proof: Note that if some edge in R belongs to a cycle of length less than $\epsilon\log n$, for $\epsilon = 1/(10\log(e^2 d/2))$, in $G_{n,d/n}$, then must exist a subgraph of $G_{n,d/n}$ that contains at most $2\epsilon\log n$ vertices and has at least two cycles.

In turn this implies that there is a set of vertices in $G_{n,d/n}$ sharing with each other a number of edges that exceeds by 1, or more, the number of vertices. The size of this set is at most $2\epsilon\log n$. We are going to show that such a set of vertices does not exist in $G_{n,d/n}$ with probability at least $1 - n^{-0.9}$. Let D be

the event that in $G_{n,d/n}$ there exists a set of r vertices which have $r+1$ edges between them. For $r \leq c \log n$ we have the following:

$$\begin{aligned} Pr[D] &\leq \sum_{r=1}^{c\log n} \binom{n}{r}\binom{\binom{r}{2}}{r+1}(d/n)^{r+1}(1-d/n)^{\binom{r}{2}-(r+1)} \\ &\leq \sum_{r=1}^{c\log n}\left(\frac{ne}{r}\right)^r\left(\frac{r^2e}{2(r+1)}\right)^{r+1}(d/n)^{r+1} \leq \frac{e\cdot d}{2n}\sum_{r=1}^{c\log n}\left(\frac{e^2d}{2}\right)^r \\ &\leq \frac{C}{n}\left(\frac{e^2d}{2}\right)^{c\log n} \end{aligned}$$

Having $c \cdot \log(e^2d/2) < 1$, the quantity in the last derivation is $o(1)$, i.e. it is of order $\Theta(n^{c\log(e^2d/2)-1})$. Thus, for ϵ as large as defined in the statement there is no connected component in G_0 which contains two cycles with probability at least $1-n^{-0.9}$

With standard probabilistic arguments we get that the expected number of cycles of length l in $G_{n,d/n}$ is less than d^l. By the linearity of expectation we conclude that the expected number of cycles of length at most $\epsilon \log n$, is at most $\Theta(n^{1/10})$. The expected number of vertices in these cycles is at most $\Theta(n^{1/10}\log n)$. All these imply, also, that the expected number of edges in R is at most $\Theta(n^{1/10}\log n)$.

Applying the Markov inequality, we get that the probability of the event that the number of edges in R is at least $n^{3/10}$ is upper bounded by $n^{-1.9/10}$. The lemma follows. ◇

Remark. Lemma A.1.3 implies that removing the edges in R, w.h.p. all the cycles of length less that $\frac{\log n}{10\log(e^2d/2)}$ become isolated.

Lemma A.1.4. *Consider the sequence of subgraphs $\mathcal{G}(G_{n,d/n})$, defined in Section A.1. W.r.t. $\mathcal{G}(G_{n,d/n})$ and when k is constant, with probability at least $1-n^{-c}$, every C_{ij}, for $0 \leq i \leq i_0$, defined in the statement of Theorem 3.2, is upper bounded by constants, $c>0$ is a constant as well.*

Proof: Considering the sequence of subgraphs $G_{i,0},\ldots G_{i,r_i}$, $u_{i,j}$ and v_{ij} are such that they are not adjacent in $G_{i,j}$ while they become adjacent in $G_{i,j+1}$. By Theorem 3.2, we have that

$$C_{ij} = C_{i,j}(G_{i,j},k) = \max_{s,t\in[k]}\left\{(Pr[X_{i,j}(u_{i,j})=s|X_{i,j}(v_{i,j})=t])^{-1}\right\}.$$

Note that C_{ij} expresses a correlation between the colour assignment of the vertices $v_{i,j}$ and $u_{i,j}$ in $PCS(G_{i,j},k)$. We will show that $C_{i,j}$ is small by comparing $Pr[X_{i,j}(u_{i,j})=s|X_{i,j}(v_{i,j})=t]$ with $Pr[X_{i,j}(u_{i,j})=s]$ and by showing that these two probability terms does not differ very much. It holds that

$$|Pr[X_{i,j}(u_{i,j})=s|X_{i,j}(v_{i,j})=t]-Pr[X_{i,j}(u_{i,j})=s]| \leq \max_{\sigma,\eta\in\Omega(G_{i,j},k)}||\mu_{ij}(\cdot|\sigma_{v_{i,j}})-\mu_{ij}(\cdot|\eta_{v_{i,j}})||_{u_{ij}}$$

Lemma follows by showing that with sufficiently large probability it holds that

$$\max_{\sigma,\eta\in\Omega(G_{i,j},k)}||\mu_{ij}(\cdot|\sigma_{v_{i,j}})-\mu_{ij}(\cdot|\eta_{v_{i,j}})||_{u_{ij}} \leq \frac{1}{100k} \qquad \text{for every } i,j. \tag{20}$$

Furthermore, there is a set of edges $\mathcal{E}$ in $G_{n,d/n}$ such that for every $e \in \mathcal{E}$ with $e = \{w_1, w_2\}$ there is set of pair of indices $I_e \times J_e$ such that for $(i,j) \in I_e \times J_e$ it holds $C_{i,j} = Pr[X_{i,j}(u_{i,j}) = s | X_{i,j}(v_{i,j}) = t]$ where either $v_{ij} = w_1$ and $u_{ij} = w_2$ or $v_{ij} = w_2$ and $u_{ij} = w_1$.

Corollary A.1.1 implies the following: For every edge $e \in \mathcal{E}$ such that the relation in (20) holds for all members of $I_e \times J_e$ if it holds for the Gibbs distribution of $PCS(G_{n,d/n} \backslash \{e\}, k)$, with an abuse of notation we use $G_{n,d/n} \backslash \{e\}$ to denote the graph $G_{n,d/n}$ without the edge $\{e\}$.

Due to the way we consider the edges in the schema the set of edges $\mathcal{E}$ contain only edges such that the smaller cycle that contains them, in $G_{n,d/n}$, is of length greater than $\epsilon \log n$, where $\epsilon = (10 \log(e^2 d/2))^{-1}$. Thus, for $e \in \mathcal{E}$ with $e = \{x, y\}$, the vertices x, y are at distance at least $\frac{\log n}{10 \log(e^2 d/2)}$ in $G_{n,d/n} \backslash \{e\}$. Thus, by Lemma A.1.2, it holds

$$E\left[\max_{\sigma,\eta \in \Omega(G_{n,d/n},k)} ||\mu(\cdot|\sigma_x) - \mu(\cdot|\eta_x)||_y\right] \leq C \cdot \frac{1}{n} \gamma^{\epsilon \log n} = C n^{-(1+\epsilon \log \gamma)}.$$

Applying the Markov inequality and for fixed C' we get

$$Pr\left[\max_{\sigma,\eta \in \Omega(G_{n,d/n},k)} ||\mu(\cdot|\sigma_x) - \mu(\cdot|\eta_x)||_y \geq \frac{1}{100k}\right] \leq C' n^{-(1+\epsilon \log \gamma)}.$$

Using Chernoff bounds it is easy to show that with probability at least $1 - n^{-\omega(n)}$ the number of edges in $G_{n,d/n}$ is at most $\frac{dn}{2}(1 + o(1))$ and consequently $|\mathcal{E}| \leq \Theta(n)$ with the same probability. Lemma follows by taking the union bound over all the pairs $x, y \in V$ such that $\{x, y\} \in \mathcal{E}$. $\diamond$

Lemma A.1.5. *Consider the sequence of subgraphs $\mathcal{G}(G_{n,d/n})$, defined in Section A.1. W.r.t. $\mathcal{G}(G_{n,d/n})$ and when $k > 2.1d$ all the probability terms $Pr[X_i(v_i) \neq X_i(u_i)]$ are lower bounded by a constant, with probability at least $1 - n^{-c}$ for c constant.*

Proof: We consider two cases for v_i and u_i: First, consider that the smallest cycle which contains both v_i and u_i is of length greater than $\frac{\log n}{10 \log(e^2 d/2)}$. Second, we consider the opposite i.e. v_i, u_i belong to a small cycle. For the first case we note that

$$|Pr[X_i(u_i) \neq X_i(v_i] - \left(1 - \frac{1}{k}\right)| \leq \max_{\sigma,\eta \in \Omega(G_i,k)} ||\mu_i(\cdot|\sigma_{v_i}) - \mu_i(\cdot|\eta_{v_i})||_{u_i}.$$

We work in the same manner as in the proof of Lemma A.1.4.

In the later case, note that we consider v_i and u_i which belong to small cycle only when we have already removed the edges in R, in G_i. Then, all the small cycles become isolated. Thus, for the second case, where the vertices belong to a small cycle, $Pr[X_i(u_i) \neq X_i(v_i]$ is trivially constant, since $k >> 2$. Lemma follows. $\diamond$

Using Lemma 2.1 and the previous lemma we get the following corollary.

Corollary A.1.2. *The log-partition function of $G_{n,d/n}$ is w.h.p. $\Theta(n)$.*

We have all the lemmas we need to show Theorem A.1.1.

Proof of Theorem A.1.1: By Theorem 3.1 the error in the estimation depends on the sum

$$\sum_{i=0}^{r-1} \frac{|Pr[X_i(v_i) \neq X_i(u_i)] - Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]|}{Pr[X_i(v_i) \neq X_i(u_i)]}.$$

According to Theorem 3.2, for $0 < i < i_0$ it holds that

$$|Pr[X_i(v_i) \neq X_i(u_i)] - Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]| \leq \sum_{j=1}^{r_i} C_{ij} \sum_{x \in \Psi_{i,j}} \max_{\sigma,\tau,\in\Omega(G_{ij},k)} ||\mu_{ij}(\cdot|\sigma_x) - \mu_{ij}(\cdot|\tau_x)||_{\Psi_i}.$$

For $i \geq i_0$ the schema makes the estimation of $Pr[X_i(v_i) \neq X_i(u_i)]$ without computations. By Corollary A.1.1 and by noting that $\mu_{i,r_i}(\cdot) = \mu_i(\cdot)$ we have that

$$|Pr[X_i(v_i) \neq X_i(u_i)] - Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]| \leq \max_j(C_{i,j}) \cdot r_i \cdot \max_{\sigma,\tau,\in\Omega(G_i,k)} ||\mu_i(\cdot|\sigma_x) - \mu_i(\cdot|\tau_x)||_{\Psi_i}$$

where x is a vertex at distance $a \log n$ from Ψ_i. Using Lemma A.1.2 and Corollary A.1.1 we get

$$E\left[\max_{\sigma,\tau\in\Omega(G_i,k)} ||\mu_i(\cdot|\sigma_x) - \mu_i(\cdot|\tau_x)||_{\Psi_i}\right] \leq C_2 n^{-(1+a\log\gamma)} \tag{21}$$

where $0 \leq a < \frac{1}{2\log d}$, γ is a fixed real greater than 1 and C_2 is a sufficiently large constant. We remind the reader that the distance between x and Ψ is $a \log n$.

The number of terms in the sequence $\mathcal{G}(G_i)$ is equal to the number of edges between $L_i(\Psi_i, a\log n)$ and $L_i(\Psi_i, a\log n + 1)$. The number of vertices of a BFS tree of $B_i(\Psi_i, a\log n)$ is dominated by a Galton-Watson tree of $a \log n$ levels, with number of offsprings per individual distributed as in $\mathcal{B}(n, d/n)$. With standard arguments (e.g. see Theorem 6 in [18]), it holds that with probability at least $1 - o(n^{-2})$, the number of vertices at level $a \log n$, where $0 \leq a \leq \frac{1}{2\log d}$, is $5n^{1/2}\log n$. Using Chernoff bounds we can show that the number of edges are $\Theta(n^{1/2}\log n)$ with probability at least $1 - o(n^{-2})$. Thus $r_i = \Theta(n^{1/2}\log n)$ with probability at least $1 - o(n^{-2})$.

Let D be the event that: a) $\max_{i,j}(C_{i,j})$ is constant, b) $\min_i\{Pr[X_i(v_i) \neq X_i(u_i)]\} \geq 0.99/k$ c) $\max_i\{r_i\} \leq \Theta(n^{1/2}\log n)$ and d) $R = n^{0.3}$. By previous lemmas we have that there is a constant $\epsilon > 0$ such that $Pr[D] \geq 1 - n^{-\epsilon}$. Combining this fact with (21) we get

$$E\left[\max_{\sigma,\tau\in\Omega(G_i,k)} ||\mu_i(\cdot|\sigma_x) - \mu_i(\cdot|\tau_x)||_{\Psi_i} \mid D\right] \leq C n^{-(1+a\log\gamma)}$$

for a sufficiently large C. Assuming that the event D holds, it also holds that

$$\frac{|Pr[X_i(v_i) \neq X_i(u_i)] - Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]|}{Pr[X_i(v_i) \neq X_i(u_i)]} = O(1) \qquad \text{for } i \geq i_0.$$

Thus, by Theorem 3.1 we get that

$$E\left[\frac{1}{n}|\log\hat{Z}(G,k) - \log Z(G,k)| \mid D\right] \leq C\left(n^{-(1/2+a\log\gamma)}\log n + n^{-1+0.3}\right)$$

Using the Markov inequality we get the following: Applying the schema on a graph $G_{n,d/n}$ where the event D holds, then with probability at least $1 - n^{-1/2}$ it holds that

$$\frac{1}{n}|\log \hat{Z}(G,k) - \log Z(G,k)| \leq \Theta(n^{-(a\log\gamma)}\log n)$$

Theorem follows by the fact that $Pr[D] \geq 1 - n^{-\epsilon}$ and Corollary A.1.2. $\diamond$

A.1.2 Proof of Theorem A.1.3

We will be based on Lemma 2.1 to show the theorem.

Consider the following sequence of subgraphs $\mathcal{G}(G_{n,d/n})$ (different than the one stated in Section A.1): The term-graph G_0 is empty. There are indices i_0 and i_1 with $i_0 \leq i_1$ while the followings hold: For $0 < i \leq i_1$, G_i contains all the edges that belong to cycles of length at most $\frac{\log n}{10\log(e^2 d/2)}$ in $G_{n,d/n}$ and only these edges. Additionally, for $0 < i \leq i_0$, G_i consists only of isolated vertices and connected components which are paths. We refer to the cycle of length less than $\frac{\log n}{10\log(e^2 d/2)}$ as "small cycles"

Working as in the proof of Lemma A.1.3 we get the following: With probability n^{-c}, for fixed $c > 0$, it holds that the number of edges in small cycles is $n^{0.3}$ and all these "small cycles" do not share vertices with each other.

By Lemma A.1.2 for $PCS(G_{n,d/n}, k)$, where $k > 2.1d$, for every $x \in V$ and $\Psi \subseteq L(x,l)$, where $0 < l < n^{1/2}$, there exists a real $\gamma > 1$ such that

$$E\left[\max_{\sigma,\eta\in\Omega(G_{n,d/n},k)} ||\mu(\cdot|\sigma_x) - \mu(\cdot|\eta_x)||_\Psi\right] \leq C \cdot \frac{|\Psi|}{n}\gamma^{-l} \tag{22}$$

where $|\Psi|$ is the number of vertices in Ψ and C is a sufficiently large real constants. By (22) and the Markov inequality we get that with probability at least $n^{-\epsilon_1}$, it holds that

$$||\mu(\cdot|\sigma_{v_i}) - \mu(\cdot|\eta_{v_i})||_{u_i} \leq n^{-\epsilon_2}$$

for $i > i_1$ where $\epsilon_1 > 0$, $\epsilon_2 > 0$ are fixed reals. From the above inequality we get that for $i > i_1$

$$Pr[X_i(v_i) \neq X_i(u_i)] = \sum_{j\in[k]} Pr[X_i(u_i) \neq j | X_i(v_i) = j] Pr[X_i(v_i) = j]$$

and

$$|Pr[X_i(u_i) \neq j | X_i(v_i) = j] - Pr[X_i(u_i) \neq j]| \leq ||\mu(\cdot|\sigma_{v_i}) - \mu(\cdot|\eta_{v_i})||_{u_i}.$$

Thus, we get that

$$Pr[X_i(v_i) \neq X_i(u_i)] = \left(1 - \frac{1}{k}\right)(1 \pm o(1))$$

Note that the above relation holds for almost all values of i in $\{1,\ldots,r\}$. E.g. there is a set of values for i, call it $A \subseteq \{1,\ldots r\}$, for which the above does not hold while $|A| = o(n)$. For $0 < i \leq i_1$ we assume that

$$Pr[X_i(v_i) \neq X_i(u_i)] = \left(1 - \frac{1}{k}\right).$$

The error that in the estimations for $0 < i \leq i_1$ is constant. Note that all $\{v_i, u_i\}$, for $0 < i \leq i_1$ belong to isolated cycles in G_i, of length less than $\frac{\log n}{(10 \log(e^2 d/2))}$. Furthermore, the error in the approximation of the log-partition is not affected by more than $o(1)$, because $i_1 = o(n)$. It is direct that it holds $Z(G_0, k) = k^n$. Theorem follows.

A.1.3 Proof of Theorem A.1.4

Let $\mathcal{G}(G)$ be the sequence of subgraphs we consider in the proof of Theorem A.1.3. It is direct that the theorem will follow by showing that we can verify that

$$||\mu(\cdot|\sigma_{v_i}) - \mu(\cdot|\eta_{v_i})||_{u_i} \leq n^{-\epsilon_1}$$

for $i > i_1$ and appropriate fixed $\epsilon_1 > 0$, in polynomial time. Note that for $\mathcal{G}(G)$ it holds that $dist(v_i, u_i) \geq \frac{\log n}{10 \log(e^2 d/2)}$, for $i > i_1$. Using Theorem 3.3 and Corollary A.1.1 we have that

$$||\mu(\cdot|\sigma_{v_i}) - \mu(\cdot|\eta_{v_i})||_{u_i} \leq 2\mathcal{P}_{\boldsymbol{q}}[\text{ } \exists \text{ path of disagreement connecting } \{v_i\} \text{ and } \{u_i\} \text{ in } G].$$

where $\mathcal{P}_{\boldsymbol{q}}$ is the the product measure defined in the statement of Theorem 3.3 w.r.t. $G_{n,d/n}$ and k. Consider the event

$$E_{v,\epsilon} = \text{“}\exists \text{ a path of disagreement that connects } v \text{ with } L(v, \epsilon \log n)\text{”}$$

for some fixed $\epsilon \geq (10 \log(e^2 d/2))$. For each pair v_i u_i define

$$a_i = \min \left\{ \frac{dist(v_i, u_i)}{\log n}, (4 \log(e^2 d/2))^{-1} \right\}.$$

Noting that, for fixed c_1, c_2 such that $c_1 > c_2$ it holds that $\mathcal{P}_{\boldsymbol{q}}[E_{v,c_1}] \leq \mathcal{P}_{\boldsymbol{q}}[E_{v,c_2}]$ we get that

$$\mathcal{P}_{\boldsymbol{q}}[\exists \text{ path of disagreement connecting } \{v_i\} \text{ and } \{u_i\} \text{ in } G] \leq \mathcal{P}_{\boldsymbol{q}}[E_{v,a_i}]$$

It remains to show that $\mathcal{P}_{\boldsymbol{q}}(E_{v_i,a_i})$, for $i \geq i_1$, can be verified that are sufficiently small in polynomial time. Let T_{v_i,a_i} be the set of all simple paths that connect v_i to $L(v_i, a_i \log n)$, it holds that

$$\mathcal{P}_{\boldsymbol{q}}[E_{v_i,a_i}] \leq \sum_{m \in T_{v_i,a_i}} \mathcal{P}_{\boldsymbol{q}}[\text{“}m \text{ is a path of disagreement”}]$$

The computation of each probability term in the r.h.s. of the above inequality can be carried out in polynomial time.

Furthermore, using Lemma 2.1 from [7] we have that for for every $v_i \in G_{n,d/n}$ the graph $B(v_i, a_i \log n)$, is tree with at most one additional edge, with probability at least $1 - n^{-0.1}$. It is direct that the number of simple paths between v and $L(v_i, a_i \log n)$ in $B(v_i, a_i \log n)$ is at most $2|L(v_i, a_i \log n)|$.

With standard arguments (e.g. see Theorem 6 in [18]), it holds that with probability at least $1 - o(n^{-2})$, $|L(v_i, a_i \log n)| = n^{0.26} \log n$. Thus, $|T_{v_i,a_i}|$ is polynomially large w.h.p.. It is direct that the above sum can be computed in polynomial-time. Theorem follows by noting that with probability at least $1 - n^{-4}$ the number of terms in $\mathcal{G}(G)$ are $\Theta(n)$.

A.2 Applications - Locally dense graphs

We consider $G = (V, E)$ to be a locally α-dense graph with bounded Δ and some $\alpha \in [0, 1]$. For a locally α-dense graph G, the cardinality of $\mathcal{BP}(v, \Lambda)$, for some $v \in V$ and $\Lambda \subseteq V$, is decreasing in the parameter of the model α. Based on this remark and considering Theorem 3.4, we show that the r.h.s. of (12) is decreasing in the distance between v, Λ as long as $k > (2 - \alpha)\Delta$.

When we consider $\mathcal{G}(G)$, note that deleting edges in G arbitrarily, we may increase the number of path between v, Λ whose vertices form an induced subgraph which is bipartite. In terms of accuracy of counting, this is a undesirable event. The sequence of subgraphs we consider below is taken so as to control the cardinality of the set $\mathcal{BP}(v_i, \Lambda)$ w.r.t. to the graph G_i. We denote this set as $\mathcal{BP}_i(v_i, \Lambda)$.

Sequence of subgraphs. Let r be the number of terms in $\mathcal{G}(G)$. We are going to describe the terms of the sequence of subgraphs in reverse order, i.e. starting from G_r and getting backwards. Clearly G_r is the same as G. Then choose some vertex of G_r, arbitrarily, let w be that vertex. In each of $G_{r-1}, G_{r-2}, \ldots$ remove only edges which are adjacent to w. When w becomes isolated, choose another vertex and continue in the same manner until all vertices become isolated.

For $0 \leq i < r$ consider the sequence of subgraphs $\mathcal{G}(G_i)$ as follows: Choose some vertex at distance t from Ψ_i, let w_i be that vertex. For $G_{i,r_i}, G_{i,r_i-1} \ldots$, remove only edges which are adjacent to w_i. Then, choose some other vertex at distance t from Ψ_i and continue in the same manner, until there are no vertices at distance t from Ψ_i.

Computing Probabilities. Set the parameter of the Counting Schema $t = c \log\log n$, for sufficiently small $c > 0$, such that the size of the connected component that contains $\{v_i, u_i\}$ in $G_{i,0}$, has at most $\Theta(\log n)$ vertices, for $0 \leq i < r$. For such component the number of colourings is polynomially large, i.e. it is at most $k^{\Theta(\log n)}$. In this case it is direct that $Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]$ can be computed by exhaustive enumeration.

It turns out that we get a $(\log n)^{-\Omega(1)}$-approximation of $\log(Z(G, k))$. At the end of this section we will see cases where the approximation of $\log(Z(G, k))$ can be improved.

In what follows, we provide the pseudocode of the counting schema for a locally α-dense graph G. The input parameter ϵ controls the quality of the approximation of $\log(Z(G, k))$. Let k be of the form $k = (1 + \beta)\Delta$ for some $\beta > 0$.

Counting Schema G

Input: G, β, t, ϵ
If $\epsilon < ((1 - \alpha)/\beta)^t$ compute $\log(Z(G, k))$ by exhaustive enumeration.
Compute the sequence of subgraphs $\mathcal{G}(G)$.
For $0 < i < r$ do

- Compute the exact value of $Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]$.
- Set $\hat{Z}(G, k) = \hat{Z}(G, k) \cdot Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]$.

End for.
Set $\hat{Z}(G, k) = \hat{Z}(G, k) \cdot k^n$.
Output $\log\left(\hat{Z}(G, k)\right)/n$.

Theorem A.2.1. *Consider $G = (V, E)$ a locally α-dense graph of bounded Δ. The Counting Schema when is applied to G with parameters $k = (1+\beta)\Delta$ and t, for β such that $1 < \alpha + \beta$, it returns an approximation of the log-partition function such that*

$$\frac{1}{n}\left|\log \hat{Z}(G,k) - \log Z(G,k)\right| \leq C\left(\frac{1-\alpha}{\beta}\right)^t$$

where C is a sufficiently large real constant.

The proof of Theorem A.2.1, makes a heavy use of Theorem 3.4, is presented in Section A.2.1.

Note that for β such that $1 < \alpha + \beta$, or equivalently, for $k > (2-\alpha)\Delta$, the quantity $\frac{1-\alpha}{\beta}$ is less than 1.

Corollary A.2.1. *It holds that $\log Z(G,k) = \Theta(n)$, for Δ, k fixed and $k > \Delta$.*

Proof: Corollary follows by Lemma A.2.2 we provide for the proof of Theorem A.2.1 ◇

Theorem 1.2 follows as a corollary from Theorem A.2.1, Corollary A.2.1 and the previous paragraph titled *Computing Probabilities*

Improved accuracy. Taking $t = c\log n$ for sufficiently small $c > 0$, the size of the connected component that contains $\{v_i, u_i\}$ in $G_{i,0}$, may become polynomially large. In this case the number of k-colourings of $G_{i,0}$ can be exponentially large and exhaustive enumeration is prohibited. However, if this component has bounded treewidth, then the *junction tree algorithm* (see [23]) can compute $Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]$ in polynomial time. In this case, Theorem A.2.1 and Corollary A.2.1 specify that we get a $n^{-\Omega(1)}$-approximation of $\log(Z(G,k))$.

The *junction tree* algorithm is a general algorithmic framework which allows calculations of marginals in settings as the one here. The detailed description of the algorithm is beyond to scope of this paper. For more details on the subject see [23]. We should, only, remark that given a *tree decomposition* of the component that contains $\{v_i, u_i\}$ in $G_{i,0}$, of *width* TR, the *junction tree* algorithm computes $Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]$ in time $O(n \cdot k^{TR})$. Note that the algorithm of Reed in [20] returns a tree decomposition of a graph G_0 of width at most TR (for fixed TR) if there is any. The time is needed is $O(N \log N)$, where N is the number of vertices of G_0. Corollary 1.1 follows.

A.2.1 Proof of Theorem A.2.1

Lemma A.2.1. *All the quantities $C_{i,j}$ defined in the statement of Theorem 3.2 are upper bounded by some constant when $k > \Delta$, and both k, Δ are fixed constants*

Proof: For any $t \in [k]$ it suffice to show that

$$\min_{s\in[k]}\{Pr[X_{i,j}(u_{i,j}) = s | X_{i,j}(v_{i,j}) = t]\} \leq C^{-1}\max_{s\in[k]}\{Pr[X_{i,j}(u_{i,j}) = s | X_{i,j}(v_{i,j}) = t]\} \tag{23}$$

where C is a sufficiently large constant. Note that in $G_{i,j}$ the minimum graph distance between $v_{i,j}$ and $u_{i,j}$ is 2. For given $t \in [k]$ let $s_M \in [k]$ be such that

$$\max_{s\in[k]} Pr[X_{i,j}(u_{i,j}) = s | X_{i,j}(v_{i,j}) = t] \leq Pr[X_{i,j}(u_{i,j}) = s_M | X_{i,j}(v_{i,j}) = t]$$

and similarly let $s_m \in [k]$ be such that

$$\min_{s\in[k]} Pr[X_{i,j}(u_{i,j}) = s | X_{i,j}(v_{i,j}) = t] \geq Pr[X_{i,j}(u_{i,j}) = s_m | X_{i,j}(v_{i,j}) = t].$$

Let $\tau \in \Omega(G_{i,j}, k)$ be such that $\tau_{v_{i,j}} = t$. For $\sigma \in \Omega(G_{i,j}, k, \tau_{v_{ij}})$ and $s \in [k]$ let $A(s, \sigma)$ be the number of colourings of $G_{i,j}$ where $u_{i,j}$ is coloured as s and the vertices in $\bar{B}_{i,j}(u_{i,j}, 2)$ are coloured as $\sigma_{\bar{B}_{i,j}(u_{i,j},2)}$. It holds that

$$Pr[X_{i,j}(u_{i,j}) = s_M | \sigma_{\bar{B}_{i,j}(u_{i,j},2)}] = \frac{A(s_M, \sigma)}{|\Omega(G_{i,j}, k, \sigma_{\bar{B}_{i,j}(u_{i,j},2)})|}$$

and

$$Pr[X_{i,j}(u_{i,j}) = s_m | \sigma_{\bar{B}_{i,j}(u_{i,j},2)}] = \frac{A(s_m, \sigma)}{|\Omega(G_{i,j}, k, \sigma_{\bar{B}_{i,j}(u_{i,j},2)})|}$$

Which implies that

$$\frac{Pr[X_{i,j}(u_{i,j}) = s_m | \sigma_{\bar{B}_{i,j}(u_{i,j},2)}]}{Pr[X_{i,j}(u_{i,j}) = s_M | \sigma_{\bar{B}_{i,j}(u_{i,j},2)}]} = \frac{A(s_m, \sigma)}{A(s_M, \sigma)}.$$

Clearly $A(s_M, \sigma)$ and $A(s_m, \sigma)$ depend on how many colourings are available for the vertices in $L_{i,j}(u_{i,j}, 1)$ given that the vertices in $\bar{B}_{ij}(u_{i,j}, 2)$ are coloured as $\sigma_{\bar{B}_{ij}(u_{i,j},2)}$ and the vertex $u_{i,j}$ is coloured as s_M and s_m, correspondingly. Since $k > \Delta$ and both k, Δ are fixed numbers, both $A(s_M, \sigma)$ and $A(s_m, \sigma)$ are bounded and non zero. This implies that there exists a sufficiently large constant C that

$$Pr[X_{i,j}(u_{i,j}) = s_M | \sigma_{\bar{B}_{i,j}(u_{i,j},2)}] \leq C \cdot Pr[X_{i,j}(u_{i,j}) = s_m | \sigma_{\bar{B}_{i,j}(u_{i,j},2)}]$$

for any $\sigma_{\bar{B}_{i,j}(u_{i,j},2)} \in \Omega(G_{i,j}, k, \tau_{v_{ij}})$. Multiplying both sides in the above inequality with $Pr[X_{i,j}(\bar{B}_{i,j}(u_{i,j}, 2)) = \sigma_{\bar{B}_{i,j}(u_{i,j},2)}]$ and taking the sum over all possible $\sigma_{\bar{B}_{i,j}(u_{i,j},2)}$ we get that

$$Pr[X_{i,j}(u_{i,j}) = s_M | X_{i,j}(v_{i,j}) = t] \leq C \cdot Pr[X_{i,j}(u_{i,j}) = s_m | X_{i,j}(v_{i,j}) = t]$$

Lemma follows. ◇

Lemma A.2.2. *For $0 \leq i < r$ it holds that $Pr[X_i(v_i) \neq X_i(u_i)]$ is lower bounded by some constant when $k > \Delta$, and both k, Δ are fixed constants.*

Proof: Very similar to the proof of Lemma A.2.1. ◇

Definition A.2.1. *For some positive integer l, let $\mathcal{BP}_i(\Psi_i, \Lambda, l) \subseteq \mathcal{BP}_i(\Psi_i, \Lambda)$ contain only the paths in $\mathcal{BP}_i(\Psi_i, \Lambda)$ which have length l.*

In the following lemma when we refer to a sequence of subgraphs $\mathcal{G}(G)$ where we delete the edges as described in Section A.2.

Lemma A.2.3. *For any graph $G = (V, E)$, consider the sequence of subgraphs $\mathcal{G}(G)$. Assume that $\mathcal{G}(G) = G_0, \ldots, G_r$. For $0 \leq i \leq r$ and $\Lambda \subseteq V$ it holds that*

$$|\mathcal{BP}_i(\Psi_i, \Lambda, l)| \leq \sum_{w \in N^r_{\Psi_i}} |\mathcal{BP}_r(w, \Lambda, l-1)|$$

where $N^r_{\Psi_i}$ is the set of neighbours of Ψ_i in G_r and l is an positive integer.

Proof: For every vertex $v \in V$, let G_{i_v} be the term of $\mathcal{G}(G)$ which has the greatest index among the terms of $\mathcal{G}(G)$ having the vertex v isolated.

Consider the permutation of the vertices in G, e.g. $v_1, \ldots, v_n$, such that, the vertex v_i is the ith vertex that becomes isolated as we look $\mathcal{G}(G)$ from the term r and backwards. Ties are resolved arbitrarily. For any vertex $v \in V$ and $\Lambda \subseteq V$ it holds that

$$|\mathcal{BP}_r(v, \Lambda, l)| \geq |\mathcal{BP}_{i_{v_1}}(v, \Lambda, l)|. \tag{24}$$

For any simple path in G_r, that start from v, ends in Λ and contains v_1, the removal of v_1 will just disconnect it. Clearly this implies that removing v_1 and all its adjacent edges from G_r, does not increase the cardinality of the paths that connect v with Λ such that their vertices form an induced subgraph which is bipartite. Repeating the above argument inductively we get that

$$|\mathcal{BP}_{i_{v_s}}(v, \Lambda, l)| \geq |\mathcal{BP}_{i_{v_{s+1}}}(v, \Lambda, l)| \tag{25}$$

for $s = 1, \ldots, n-1$. Note that $i_{v_s} \geq i_{v_{s+1}}$. Consider, now, the graph G_j, in G(G), for some $j \in [i_{v_{s+1}}, \ldots i_{v_s} - 1]$. It holds

$$|\mathcal{BP}_j(\Psi_j, \Lambda, l)| \leq \sum_{w \in N^{i_{v_s}}_{\Psi_j}} |\mathcal{BP}_{i_{v_s}}(w, \Lambda, l-1)|.$$

Note that in the above inequality, each summand in r.h.s. allows to use the vertices in Ψ_j. Lemma follow by noting that (25) implies

$$|\mathcal{BP}_j(\Psi_j, \Lambda, l)| \leq \sum_{w \in N^r_{\Psi_j}} |\mathcal{BP}_r(w, \Lambda, l-1)|.$$

$\diamond$

Proof of Theorem A.2.1: Let $M = (x_1 \ldots, x_l)$ such that $M \in \mathcal{BP}_r(\{x_1\}, \{x_l\}, l-1)$. Assume that we want to add in M a vertex x_{l+1} such that the new path $M' = (x_1, \ldots, x_{l+1})$ to belong to $\mathcal{BP}_r(\{x_1\}, \{x_{l+1}\}, l)$. M can be extended only using vertices that are not adjacent to x_{l-1}. Since G_r is locally α-dense, from x_l there are only, at most, $(1-\alpha)\Delta$ available vertices.

Note that if the path could be extended to a vertex in the neighbourhood of x_{l-1}, then x_{l-1} x_l and the new vertex would form a triangle. This would imply that $M' \notin \mathcal{BP}_r(\{x_1\}, \{x_{l+1}\}, l)$. Consequently, the number of paths that start from $v \in V$ have length l and their vertices form a induced subgraph of G_r which is bipartite is, at most, $\Delta\left((1-\alpha)\Delta\right)^{l-1}$.

Theorems 3.2 and 3.4 state that

$$|Pr[X_i(v_i) \neq X_i(u_i)] - Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]| \leq \sum_{j=1}^{r_i} C_{ij} \sum_{x \in \Psi_{i,j}} \sum_{M \in \mathcal{BP}_{ij}(x, \Psi_i)} E_{\boldsymbol{q}}[\mathcal{I}_M]. \tag{26}$$

The expectation in the r.h.s. is taken w.r.t. to product measure $\mathcal{P}_{\boldsymbol{q}}$ specified by G and k. Let $\mathcal{P}'_{\boldsymbol{q}}$ be the product measure specified by G_i for some i and k. It is direct that $\mathcal{P}_{\boldsymbol{q}}$ dominates stochastically $\mathcal{P}'_{\boldsymbol{q}}$, in the sense that some vertex is more probable to be disagreeing under the first measure. Consider the quantity

$$BP(l) = \sum_{j=0}^{r_i} |\mathcal{BP}_{i,j}(\Psi_{ij}, \Psi_i, l)|.$$

Using Lemma A.2.3, we get that

$$|\mathcal{BP}_{i,j}(\Psi_{i,j}, \Psi_i, l)| \leq \sum_{w \in N^i_{\Psi_{ij}}} |\mathcal{BP}_i(w, \Psi_i, l-1)|$$

Let $L = \cup_{l=t-2}^{t+2} L_r(\Psi_i, l)$, it holds that

$$BP(l) \leq \sum_{w \in L} |\mathcal{BP}_i(w, \Psi_i, l-1)|$$

Noting that $\mathcal{BP}_i(w, \Psi_i) = \mathcal{BP}_i(\Psi_i, w)$ and applying Lemma A.2.3 we get that

$$|\mathcal{BP}_i(\Psi_i, w, l-1)| \leq \sum_{y \in N^r_{\Psi_i}} |\mathcal{BP}_r(y, w, l-2)|.$$

Thus, for $l \geq t$ (t is a parameter of the Counting Schema) we get that

$$BP(l) \leq \sum_{w \in L} \sum_{x \in N_{\Psi_i}} |\mathcal{BP}_r(w, x, l-2)| \leq 2\Delta^2 \left((1-\alpha)\Delta\right)^{l-2}.$$

We can write (26) as follows:

$$|Pr[X_i(v_i) \neq X_i(u_i)] - Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]| \leq \max_j C_{ij} \sum_{l \geq t} BP(l) \left(\frac{1}{k-\Delta}\right)^l.$$

For $k = (1+\beta)\Delta$ we have

$$|Pr[X_i(v_i) \neq X_i(u_i)] - Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]| \leq \frac{\max_j C_{ij}}{\beta^2} \sum_{l \geq t} \left(\frac{1-\alpha}{\beta}\right)^{l-2}.$$

Taking $1 < \alpha + \beta$ the ratio $(1-\alpha)/\beta$ is smaller than 1. In this case we get:

$$\begin{aligned} |Pr[X_i(v_i) \neq X_i(u_i)] - Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]| &\leq \tfrac{\max_j C_{ij}}{\beta^2} \textstyle\sum_{l \geq t} \left(\frac{1-\alpha}{\beta}\right)^{l-2} \\ &\leq \tfrac{\max C_{ij}}{\beta^2} \tfrac{1}{1-\frac{1-\alpha}{b}} \cdot \left(\tfrac{1-\alpha}{b}\right)^{l-2} \end{aligned}$$

Plugging in Lemma A.2.2 and Lemma A.2.1 we get that for $0 \leq i \leq r$

$$\frac{|Pr[X_i(v_i) \neq X_i(u_i)] - Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]|}{Pr[X_i(v_i) \neq X_i(u_i)]} \leq C \left(\frac{1-\alpha}{\beta}\right)^l.$$

where C is a sufficiently large constant. Theorem follows by using Theorem 3.1. $\diamond$

A.3 Remarks on Locally α-dense graphs

The following lemma provides a criterion under which a graph is locally α-dense. It is not too confining and allows a rich family of locally dense graphs.

Lemma A.3.1. *Consider the graph $G = (V, E)$ with maximum degree Δ. For each $v \in V$ let $Q_v = \{Q_1, Q_2, \ldots, Q_s\}$ denote the set of all maximal cliques that v belongs. If for every $v \in V$ and for every $i \in [0, \ldots, |Q_v|]$ it holds*

$$\sum_{Q_j \in Q_v \setminus \{Q_i\}} |Q_j \setminus Q_i| \leq \lfloor (1-\alpha)\Delta \rfloor \tag{27}$$

for some real $\alpha \in [0, 1]$, then G is locally α-dense.

Proof: Consider the graph $G = (V, E)$ and some $\{v, u\} \in E$. Note that the neighbours of v that are adjacent to u are only those that belong to the union of the maximal cliques that contain both v and u. Assume that one of the cliques that u belongs is Q_i. Then the summation in (27) gives an upper bound to the number of the neighbours of v that are not adjacent to u. This bound is exact if and only if, among the cliques in Q_v, u belongs only to Q_i, otherwise it is an overestimate. Clearly, if for every v and i this sum is less than $\lfloor (1-\alpha)\Delta \rfloor$, then G is locally α-dense. $\diamond$

It is easy to check using the previous lemma that if G is a locally α-dense, there can be vertex v having a *triangle free neighbourhood* as long as its degree of v and the degrees of its adjacent vertices are at most $(1-\alpha)\Delta$. Also, note that if the inequality (27) is not true, then G may still be locally α-dense.

A.4 Bounds for spatial correlation decay

In this section we present the proofs of Theorem 3.3 and Theorem 3.4, stated in Section 3.2.

For some finite graph $G = (V, E)$ and some integer k such that $\Omega(G, k) > 0$, let $\mu(\cdot)$ be the Gibbs distribution specified by $PCS(G, k)$. For $x \in V$, $\Lambda \subseteq V$ and $\sigma, \eta \in \Omega(G, k)$, we are interested in deriving upper bounds for following quantity

$$||\mu(\cdot|\sigma_x) - \mu(\cdot|\eta_x)||_\Lambda. \tag{28}$$

Towards bounding the quantity in (28) we introduce two random variables X^σ, $X^\eta \in [k]^V$ distributed as in $\mu(\cdot|\sigma_x)$ and $\mu(\cdot|\eta_x)$, correspondingly. For any $\Psi \subseteq V$ let $X^\sigma(\Psi) \in [k]^\Psi$ ($X^\eta(\Psi) \in [k]^\Psi$) be the random variable such that if $X^\sigma = \xi$ ($X^\eta = \xi$), then $X^\sigma(\Psi) = \xi_\Psi$ ($X^\eta(\Psi) = \xi_\Psi$). We couple X^σ and X^η and use the Coupling Lemma (see [2]) to bound (28) as follows:

$$||\mu(\cdot|\sigma_x) - \mu(\cdot|\eta_x)||_\Lambda \leq Pr[X^\sigma(\Lambda) \neq X^\eta(\Lambda) \text{ in the coupling}].$$

We provide a upper bound for the probability of the event "$X^\sigma(\Lambda) \neq X^\eta(\Lambda)$" in the coupling, in terms of k and the degrees of the vertices in G by using "*disagreement percolation*", [5].

The identity coupling of X^σ, X^η is precluded due to the disagreement at x. The "disagreement percolation" coupling construction assembles the coupling (we have provide at the first place) in a stepwise fashion working away from x.

In our case it may be viewed as follows. The disagreement propagates into G along paths from x. A disagreement at vertex $y \in V$ at graph distance l from x propagates to a neighbour z at distance $l+1$ if $X^{\sigma}(z) \neq X^{\eta}(z)$. A *path of disagreement* in G is any path which has all its vertices disagreeing.

The disagreement percolation is dominated by an independent process [14]. As we will show, the disagreement propagates along a self-avoiding path $\mathcal{M}$ (that starts from x) of length l, with probability at most $\prod_{v \in \mathcal{M} \setminus \{x\}} \frac{1}{k-\Delta_v}$, or $\left(\frac{1}{k-\Delta}\right)^l$.

In the following subsection we describe the coupling of X^{σ} and X^{η}. Afterwards we present the proofs of Theorem 3.3 and Theorem 3.4.

A.4.1 The coupling for the comparison

For coupling X^{σ} and X^{η}, we exploit the bijection between $\Omega(G,k,\sigma_x)$ and $\Omega(G,k,\eta_x)$, provided in the following lemma.

Lemma A.4.1. *There is a bijection $T : \Omega(G,k,\sigma_v) \to \Omega(G,k,\eta_x)$ such that for every $\xi \in \Omega(G,k,\sigma_x)$, the set $V_\xi = \{v \in V | \xi_v \neq (T\xi)_v\}$ has the following properties:*

- *for every $v \in V_\xi$ it holds $(\xi)_v, (T\xi)_v \in \{\sigma_x, \eta_x\}$*
- *the graph $G_\xi = (V_\xi, E_\xi)$, induced subgraph of G, is* connected *and* bipartite.

Proof: The proof of this lemma is based on describing the actual bijection T. More specifically, we demonstrate how T maps a given member of $\Omega(G,k,\sigma_x)$ to a specific member of $\Omega(G,k,\eta_x)$. Then we show that the mapping T is actually a bijection between the sets $\Omega(G,k,\sigma_x)$ and $\Omega(G,k,\eta_x)$ while the set V_ξ, as defined in the statement, has all the desired properties.

For the colouring $\xi \in \Omega(G,k,\sigma_x)$, let $G_\xi = (V_\xi, E_\xi)$ be the maximal, connected induced subgraph of G such that $x \in V_\xi$ and $\forall v \in V_\xi$ it holds $\xi_v \in \{\sigma_x, \eta_x\}$. It is easy for someone to see that G_ξ is a bipartite graph (in the extreme case where $V_\xi = \{x\}$ we consider G_ξ bipartite too). The maximality of G_ξ implies that if $\partial V_\xi = \{v \in V \setminus V_\xi | \{v,u\} \in E \text{ for } u \in V_\xi\}$, then $\forall v \in \partial V_\xi$ it holds $\xi_v \notin \{\sigma_x, \eta_u\}$. It is crucial that the graph G_ξ is connected.

Clearly ξ specifies a proper 2-colouring for G_ξ that uses only the colours σ_x and η_x. In particular, let $p_1, p_2 \subseteq V_\xi$ be the two parts of G_ξ and w.l.o.g. assume that x belongs to p_1. Then, ξ assigns to all the vertices in p_1 the colour σ_x^i and to all the vertices in p_2 the colour η_x.

We take T such that for every vertex $v \in V \setminus V_\xi$ to hold $(T\xi)_v = \xi_v$. For the remaining vertices, i.e. those that belong to G_ξ, the mapping T swaps the colour assignments of the two parts of G_ξ. More specifically, if ξ specifies for all the vertices in each of the parts p_1 and p_2, of the graph G_ξ, colour assignment σ_x and η_x correspondingly, then $T\xi$ specifies for all the vertices in each of the parts p_1 and p_2 colour assignments η_x and σ_x, correspondingly.

Lemma follows by showing that T is a bijection between $\Omega(G,k,\sigma_x)$ and $\Omega(G,k,\eta_x)$.

Claim A.4.1 *For every $\xi \in \Omega(G,k,\sigma_x)$ it holds that $(T\xi) \in \Omega(G,k,\eta_x)$.*

[14] It is dominated by a Bernoulli site percolation.

Proof: It is direct that $(T\xi)_x = \eta_x$. It remains to show that $T\xi$ is a proper colouring of G.

If $T\xi$ is a non proper colouring, then we will have, at least, two adjacent vertices (somewhere in G) having the same colour assignment. The swap of colour assignments that take place, when we apply T on ξ, involves only vertices in V_ξ. Thus if $(T\xi)$ is a non proper colouring, then the adjacent vertices with the same colour assignment can be either all in V_ξ or some in V_ξ and the rest in ∂V_ξ.

It is direct that swapping the colour assignments of the two parts of G_ξ, as these are specified by ξ, leads to a proper colouring of G_ξ. Thus, in $T\xi$ no adjacent vertices in G_ξ can have the same colour assignment. Also, this swap of colourings cannot lead some vertex in V_ξ to have the same colour assignment with some vertex in ∂V_ξ. Due to maximality of G_ξ, the colouring ξ does not specify colour assignment that uses the colours σ_x and η_x for any vertex in ∂V_ξ. Thus, the swap in the colour assignments that take place when T is applied on ξ does not lead to a non proper colouring. The claim follows. $\diamond$

The next claim shows that every member of $\Omega(G, k, \eta_x)$ can be derived by applying T to an appropriate member of $\Omega(G, k, \sigma_x)$, i.e. Claim A.4.2 shows that T is a surjective mapping.

Claim A.4.2 *The range of T is $\Omega(G, k, \eta_x)$.*

Proof: Let ξ' be any member of $\Omega(G, k, \eta_x)$. We are going to show that there exists $\xi \in \Omega(G, k, \sigma_x)$ such that $T\xi = \xi'$.

For the colouring ξ', let $G_{\xi'} = (V_{\xi'}, E_{\xi'})$ be the maximal, connected bipartite subgraph of G such that $x \in V_{\xi'}$ and $\forall v \in V_{\xi'}$ it holds $\xi'_v \in \{\sigma_x, \eta_x\}$ (in the extreme case where $V_{\xi'} = \{x\}$ we consider $G_{\xi'}$ a bipartite graph).

The colouring ξ' specifies a proper 2-colouring for $G_{\xi'}$ that uses only the colours σ_x and η_x. Let $p_1, p_2 \subseteq V_\xi$ be the two parts of $G_{\xi'}$ and w.l.o.g. assume that ξ' assigns to all the vertices in p_1 the colour η_x and to all the vertices in p_2 the colour σ_x.

Consider the colouring ξ which is derived by ξ' by swapping the colour assignments of the two parts of $G_{\xi'}$ while $\xi_v = \xi'_v$ for $v \in V \backslash V_{\xi'}$. With arguments similar to those in the proof of Claim A.4.1 we can see that $\xi \in \Omega(G, k, \sigma_x)$. The claim follows by noting that $T\xi = \xi'$. $\diamond$

In the following claim we show that T is one-to-one.

Claim A.4.3 *The mapping T is one-to-one.*

Proof: Assume that there are two colourings $\xi^1, \xi^2 \in \Omega(G, k, \sigma_x)$ such that $T\xi^1 = T\xi^2 = \xi^3$. We are going to show that it should hold $\xi^1 = \xi^2$.

Assume the contrary, i.e. $\xi^1 \neq \xi^2$. We consider the graphs G_{ξ^1} G_{ξ^2} and G_{ξ^3}, as in the proofs of the two previous claims. By the proofs of two previous claims we know that the graphs G_{ξ^1}, G_{ξ^2} and G_{ξ^3} are isomorphic.

We conclude that the colourings ξ^1 and ξ^2 should differ only on the colour assignment of the vertices in the graph G_{ξ^1}. We remind the reader that this graph is a connected bipartite graph with ξ^1 and ξ^2 specifying proper 2-colourings for G_{ξ^1} which both using the colours $\{\sigma_x, \eta_x\}$.

By definition, the 2-colouring for G_{ξ^1} that ξ^1 specifies is different than that of $T\xi^1$. The same holds for colouring of ξ^2 and $T\xi^2$. Since $T\xi^1 = T\xi^2$ we should

have three different proper 2-colourings for G_{ξ^1}. Noting that there exist only two proper 2-colourings for G_{ξ^1} we have a contradiction. The claim follows. $\diamond$

Since the mapping $T : \Omega(G, k, \sigma_x) \to \Omega(G, k, \eta_x)$ is surjective (Claim A.4.2) and one-to-one (Claim A.4.3), it is a bijection. The lemma is proved. $\diamond$

Lemma A.4.2. *There exists a coupling of X^σ with X^η such that*

$$X^\eta = TX^\sigma$$

where T is the bijection defined in the statement of Lemma A.4.1.

Proof: The existence of the bijection T implies that $|\Omega(G, k, \sigma_x)| = |\Omega(G_i, k, \eta_x)|$. Thus $\forall \xi \in \Omega(G, k, \sigma_x)$ it holds that

$$\mu_i\left(\xi|\sigma_x\right) = \mu_i\left((T\xi)|\eta_x\right) = \frac{1}{|\Omega(G, k, \sigma_x)|}.$$

This implies that $Pr[X^\sigma = \xi] = Pr[X^\eta = T\xi]$, $\forall \xi \in \Omega(G, k, \sigma_x)$. The lemma follows by noting that

$$\left(\sum_{\xi \in \Omega(G,k,\sigma_x)} Pr[X^\sigma = \xi]\right) = 1 \qquad \text{and} \qquad \left(\sum_{\xi \in \Omega(G,k,\sigma_x)} Pr[X^\eta = (T\xi)]\right) = 1.$$

$\diamond$

Definition A.4.1. *Let $G^d(V^d, E^d)$ denote the probability space over all the induced subgraphs of G such that $V^d = \{u \in V | X^\sigma(u) \neq X^\eta(u)$ in the coupling of Lemma A.4.2$\}$.*

The definition of the bijection T implies G^d is, always, connected and bipartite.

Theorem A.4.1. *With the above definitions and $\hat{Pr}[E]$ denote the probability of the event $E \subseteq [k]^V \times [k]^V$ to occur in the coupling defined in Lemma A.4.2, it holds that*

$$||\mu(\cdot|\sigma_x) - \mu(\cdot|\eta_x)||_\Lambda \leq \hat{Pr}[V^d \cap \Lambda \neq \emptyset] \tag{29}$$

Proof: Theorem follows by the Coupling Lemma (see in [2]). $\diamond$

We should remark that Lemma A.4.1, Lemma A.4.2 and Theorem A.4.1 not only do relate the quantity $||\mu(\cdot|\sigma_\Psi) - \mu(\cdot|\eta_\Psi)||_\Lambda$ with the probability of having a sufficiently "large" disagreement, but they give extra information regarding its "shape" and its "colouring" as well. More specifically, the vertices in the disagreement form a connected bipartite graph and their colour assignments use only σ_x and η_x. We close the presentation with a lemma which is going to be very useful in the following sections, where we apply the disagreement percolation.

Lemma A.4.3. *For every $u \in V \backslash \{x\}$, let N_u be the set that contains all the vertices which are adjacent to the vertex u in G. Also, let $\mathcal{B}_u \subseteq [k]^{N_u} \times [k]^{N_u}$ be defined such that*

$$\mathcal{B}_u = \{e \in [k]^{N_u} \times [k]^{N_u} | \hat{Pr}[e] \neq 0\}.$$

If $k > \Delta$, then it holds that

$$\max_{E \in \mathcal{B}_u} \hat{Pr}[X^\sigma(u) \neq X^\eta(u)|E] \leq \frac{1}{k - \Delta_u}$$

where Δ_u is the degree of vertex u in G.

Proof: Lemma A.4.1 and Lemma A.4.2 suggest that given the value of X^σ, the random set V^d is specified with probability 1. This implies that

$$\max_{E\in\mathcal{B}_u} \hat{Pr}[X^\sigma(u) \neq X^\eta(u)|E] = \max_{\xi\in\Omega(G,k)} \hat{Pr}[X^\sigma(v) \neq X^\eta(v)|X^\sigma(V\backslash\{v\}) = \xi_{V\backslash\{v\}}].$$

We remind the reader that the coupling of Lemma A.4.2 is such that G^d is coloured using only the colours σ_x and η_x.

There are two necessary conditions for some vertex $v \in V\backslash\{x\}$ to be in V^d. The first one is that some vertex in N_u should, also, belong to V^d. This is due to the fact that G^d is connected. The second is the following one: Assume that $w_1 \in N_u$ and $w_1 \in V^d$. If there exists $w_2 \in N_u\backslash\{w_1\}$ and $X^\sigma(w_2) \in \{\sigma_x, \eta_x\}$, then it should hold $X^\sigma(w_1) = X^\sigma(w_2)$. This should hold since G_d is bipartite.

Considering the two previous conditions the worst case of $X^\sigma(N_u)$ is the following: At least one vertex in N_u belongs to V^d, call this vertex w. No vertex in N_u uses the colour $\{\sigma_x, \eta_x\}\backslash\{X^\sigma(w)\}$. $X^\sigma(N_u)$ is such that the number of different colour that are used is equal to $|N_u|$. In that case the probability of u to belong to V^d is $\frac{1}{k-\Delta_u}$. The lemma follows. ◇

Lemma A.4.3 assumes that $k > \Delta$, otherwise it holds

$$\max_{E\in\mathcal{B}_u} \hat{Pr}[X^\sigma(u) \neq X^\eta(u)|E] \leq 1.$$

A.4.2 Proof of Theorem 3.3

By Theorem 1 and Corollary 1.1 in [5], and Lemma A.4.3 we get that

$$||\mu(\cdot|\sigma_x) - \mu(\cdot|\eta_x)||_\Lambda \leq \mathcal{P}_{\boldsymbol{q}}[\exists \text{ path of disagreement between } \{x\} \text{ and a vertex in } \Lambda] \tag{30}$$

where $\boldsymbol{q}$ is as defined in the statement of the Theorem 3.3.

We have to remark here that the coupling on which the disagreement percolation is based, has the following property: Let t be the minimum integer such that there is no path of disagreement connecting x to $L(x,t)$. Then, our coupling specifies that no vertex in $L(x,t')$, for $t' \geq t$ can be disagreeing. This is a crucial property of our coupling, since otherwise we could not apply the disagreement percolation technique (see [12]).

A.4.3 Proof of Theorem 3.4

In Theorem 3.3 it is not taken into consideration that the disagreeing vertices of the coupling between X^σ and X^η should, always, form an induced subgraph of G which is bipartite. Plugging in this property of the disagreement we improve the bound in (11) by restricting to paths of disagreement whose vertices form an induced subgraph of G which is bipartite.

For the definition of the set $\mathcal{BP}(x,\Lambda)$ and the variables $\mathcal{I}_M$ the reader should see in Section 3.2 (before the statement of Theorem 3.4).

Proof of Theorem 3.4: Consider the vertices of the path $M \in \mathcal{BP}(x,\Lambda)$ and let $(x, x_1, \ldots, x_l)$ be the order in which we see the vertices in M as we traverse it from $\{x\}$ to Λ. Let

$$P_{x_j} = \hat{Pr}[x_j \text{ is disagreeing}|x_t \text{ is disagreeing } 0 < t < j].$$

Clearly it holds that $\hat{Pr}[\mathcal{I}_M = 1] = \prod_{j=1}^{l} P_{x_j}$. By Lemma A.4.3 it holds that $P_{x_j} \leq \frac{1}{k-\Delta_{x_j}}$, thus

$$\hat{Pr}[\mathcal{I}_M] \leq \prod_{j=1}^{l} \frac{1}{k-\Delta_{x_j}}.$$

By Theorem A.4.1 we have that

$$||\mu(\cdot|\sigma_x) - \mu(\cdot|\eta_x)||_\Lambda \leq \hat{Pr}\left[\cup_{M \in \mathcal{BP}(x,\Lambda)} \mathcal{I}_M\right].$$

Theorem follows by applying the union bound. ◇

A.5 Proofs

A.5.1 Proof of Theorem 3.1

Let

$$err_i = |Pr[X_i(v_i) \neq X_i(u_i)] - Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]| \qquad \text{for } 0 \leq i \leq r-1.$$

$$\begin{aligned}
\log \hat{Z}(G,k) &= \textstyle\sum_{i=0}^{r-1} \log(P[X_{i,0}(v_i) \neq X_{i,0}(u_i)]) + \log Z(G_0,k) \\
&\leq \textstyle\sum_{i=0}^{r-1} \log\left(P[X_i(v_i) \neq X_i(u_i)] + err_i\right) + \log Z(G_0,k) \\
&\leq \textstyle\sum_{i=0}^{r-1} \log\left(P[X_i(v_i) \neq X_i(u_i)]\right) + \sum_{i=0}^{r-1} \log\left(1 + \frac{err_i}{P[X_i(v_i) \neq X_i(u_i)]}\right) + \log Z(G_0,k) \\
&\leq \log Z(G,k) + \textstyle\sum_{i=0}^{r-1} \log\left(1 + \frac{err_i}{P[X_i(v_i) \neq X_i(u_i)]}\right) \\
&\leq \log Z(G,k) + \textstyle\sum_{i=0}^{r-1} \frac{err_i}{P[X_i(v_i) \neq X_i(u_i)]}
\end{aligned}$$

The final derivation follows by the the fact that $\log(x)$ is an increasing function (since the base is of the logarithm is $e > 1$) and by the inequality $1 + x \leq e^x$, for any x. Similarly we get the lower bound for $\log(\hat{Z}(G,k))$. Theorem follows.

A.5.2 Proof of Theorem 3.2

Theorem 3.2 follows as a corollary of two following lemmas.

Lemma A.5.1. *It holds that*

$$|Pr[X_i(v_i) \neq X_i(u_i)] - Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]| \leq \sum_{j=0}^{r_i-1} ||\mu_{i,j}(\cdot) - \mu_{i,j+1}(\cdot)||_{\Psi_i}.$$

Proof: Clearly the Gibbs distributions μ_i and μ_{i,r_i} are the same.

$$|Pr[X_i(v_i) \neq X_i(u_i)] - Pr[X_{i,0}(v_i) \neq X_{i,0}(u_i)]| \leq \max_{A \subseteq [k]^{\Psi_i}} |\mu_{i,0}(A) - \mu_i(A)| \leq ||\mu_{i,0}(\cdot) - \mu_{i,r_i}(\cdot)||_{\Psi_i}$$

By the triangle inequality we get that $||\mu_{i,0} - \mu_{i,r_i}||_{\Psi_i} \leq \sum_{j=0}^{r_i-1} ||\mu_{i,j}(\cdot) - \mu_{i,j+1}(\cdot)||_{\Psi_i}$
◇

Lemma A.5.2. *Let Λ be any subset of vertices of $G_{i,j}$ that does not contain v_i and u_i. It holds that*

$$||\mu_{i,j} - \mu_{i,j+1}||_\Lambda \leq C_{i,j} \cdot \sum_{x \in \Psi_{i,j}} \max_{\sigma^1, \sigma^2 \in \Omega(G_{i,j},k)} ||\mu_{i,j}(\cdot|\sigma^1_x) - \mu_{i,j}(\cdot|\sigma^2_x)||_\Lambda \quad (31)$$

where $C_{ij} = C_{i,j}(G_{i,j}, k) = \max_{s,t \in [k]} \left\{ (Pr[X_{i,j}(u_{i,j}) = s | X_{i,j}(v_{i,j}) = t])^{-1} \right\}$.

Proof: With relatively simple argument we can get that

$$||\mu_{i,j}(\cdot) - \mu_{i,j+1}(\cdot)||_\Lambda \leq ||\mu_{i,j}(\cdot) - \mu_{i,j+1}(\cdot|\tau_{\Psi_{i,j}})||_\Lambda$$

where $\tau \in \Omega(G_{i,j+1}, k)$ is taken so as $\tau_{\Psi_{i,j}}$ to maximize the r.h.s. in the above inequality. By the fact that $\Omega(G_{i,j+1}, k) \subseteq \Omega(G_{i,j}, k)$ and by the conditional independence, it holds that $\mu_{i,j+1}(\cdot|\tau_{\Psi_{i,j}}) = \mu_{i,j}(\cdot|\tau_{\Psi_{i,j}})$, thus

$$||\mu_{i,j} - \mu_{i,j+1}||_\Lambda \leq ||\mu_{i,j}(\cdot) - \mu_{i,j}(\cdot|\tau_{\Psi_{i,j}})||_\Lambda.$$

By definition (see (3)), there exists a set $\mathcal{A} \subseteq [k]^\Lambda$ such that $\mu_{i,j}(\mathcal{A}) > 0$[15] and

$$||\mu_{i,j}(\cdot) - \mu_{i,j}(\cdot|\tau_{\Psi_{i,j}})||_\Lambda = |\mu_{i,j}(\mathcal{A}) - \mu_{i,j}(\mathcal{A}|\tau_{\Psi_{i,j}})|.$$

It holds that

$$\frac{\mu_{i,j}(\mathcal{A}|\tau_{\Psi_{i,j}})}{\mu_{i,j}(\mathcal{A})} = \frac{\mu_{i,j}(\tau_{\Psi_{i,j}}|\mathcal{A})}{\mu_{i,j}(\tau_{\Psi_{i,j}})}$$

which implies that

$$\begin{aligned} |\mu_{i,j}(\mathcal{A}|\tau_{\Psi_{i,j}}) - \mu_{i,j}(\mathcal{A})| &= \frac{\mu_{i,j}(\mathcal{A})}{\mu_{i,j}(\tau_{\Psi_{i,j}})} \cdot |\mu_{i,j}(\tau_{\Psi_{i,j}}|\mathcal{A}) - \mu_{i,j}(\tau_{\Psi_{i,j}})| \\ &\leq \frac{\mu_{i,j}(\mathcal{A})}{\mu_{i,j}(\tau_{\Psi_{i,j}})} \cdot ||\mu_{i,j}(\cdot|\mathcal{A}) - \mu_{i,j}(\cdot)||_{\Psi_{i,j}} \\ &\leq \frac{\mu_{i,j}(\mathcal{A})}{\mu_{i,j}(\tau_{\Psi_{i,j}})} \cdot \sum_{x \in \Psi_{i,j}} ||\mu_{i,j}(\cdot|\mathcal{A}) - \mu_{i,j}(\cdot)||_x \end{aligned} \quad (32)$$

The inequality in the second line holds since

$$|\mu_{i,j}(\tau_{\Psi_{i,j}}|\mathcal{A}) - \mu_{i,j}(\tau_{\Psi_{i,j}})| \leq ||\mu_{i,j}(\cdot|\mathcal{A}) - \mu_{i,j}(\cdot)||_{\Psi_{i,j}}.$$

The inequality in the third line holds by the fact that

$$||\mu_{i,j}(\cdot|\mathcal{A}) - \mu_{i,j}(\cdot)||_{\Psi_{i,j}} \leq \sum_{x \in \Psi_{i,j}} ||\mu_{i,j}(\cdot|\mathcal{A}) - \mu_{i,j}(\cdot)||_x.$$

(See Section A.5.3 Proposition A.5.1 for a proof of the above inequality, which is based on coupling.) For every $x \in \Psi_{i,j}$, there exists $\hat{\mathcal{A}} \subset [k]^x$ such that $\mu_{i,j}(\hat{\mathcal{A}}) > 0$ and

$$||\mu_{i,j}(\cdot|\mathcal{A}) - \mu_{i,j}(\cdot)||_x = |\mu_{i,j}(\hat{A}|\mathcal{A}) - \mu_{i,j}(\hat{A})|.$$

It holds that

$$\frac{\mu_{i,j}(\hat{\mathcal{A}}|\mathcal{A})}{\mu_{i,j}(\hat{\mathcal{A}})} = \frac{\mu_{i,j}(\mathcal{A}|\hat{\mathcal{A}})}{\mu_{i,j}(\mathcal{A})}$$

[15] It is direct that if $\mu_{i,j}(\mathcal{A}) = 0$, then $\mu_{i,j}(\mathcal{A}|\tau_{\Psi_{i,j}}) = 0$.

which implies that

$$|\mu_{i,j}(\hat{\mathcal{A}}|\mathcal{A}) - \mu_{i,j}(\hat{\mathcal{A}})| = \frac{\mu_{i,j}(\hat{\mathcal{A}})}{\mu_{i,j}(\mathcal{A})}|\mu_{i,j}(\mathcal{A}|\hat{\mathcal{A}}) - \mu_{i,j}(\mathcal{A})| = \frac{\mu_{i,j}(\hat{\mathcal{A}})}{\mu_{i,j}(\mathcal{A})}|\mu_{i,j}(\mathcal{A}|\hat{\mathcal{A}}) - \mu_{i,j}(\mathcal{A})|$$

It is direct that

$$|\mu_{i,j}(\mathcal{A}|\hat{\mathcal{A}}) - \mu_{i,j}(\mathcal{A})| \leq \max_{\sigma^1,\sigma^2 \in \Omega(G_{i,j},k)} ||\mu_{i,j}(\cdot|\sigma^1_x) - \mu_{i,j}(\cdot|\sigma^2_x)||_{\Lambda}.$$

Thus, we get that

$$||\mu_{i,j}(\cdot) - \mu_{i,j}(\cdot|\tau_{\Psi_{i,j}})||_{\Lambda} \leq \frac{1/k}{\mu_{i,j}(\tau_{\Psi_{i,j}})} \cdot \sum_{x \in \Psi_{i,j}} \max_{\sigma^1,\sigma^2 \in \Omega(G_{i,j},k)} ||\mu_{i,j}(\cdot|\sigma^1_x) - \mu_{i,j}(\cdot|\sigma^2_x)||_{\Lambda}$$

After some trivial derivations the lemma follows. ◇

A.5.3 Properties of Total variation distance

Proposition A.5.1. *For any $\sigma \in \Omega(G_{i,j}, k)$, and any $\Lambda \subseteq V$ it holds that*

$$||\mu_{i,j}(\cdot) - \mu_{i,j}(\cdot|\sigma_{\Lambda})||_{\Psi_{i,j}} \leq \sum_{x \in \Psi_{i,j}} ||\mu_{i,j}(\cdot) - \mu_{i,j}(\cdot|\sigma_{\Lambda})||_x$$

Proof: The convenience we abbreviate $G_{i,j}$, $\mu_{i,j}(\cdot)$ and $\Psi_{i,j}$ to G μ and Ψ, correspondingly. The proof is going to be made by coupling. Consider two systems $S_1 = PCS(G, k, \sigma_{\Lambda})$ and $S_2 = PCS(G, k)$. Let v, u be the two vertices that belong to Ψ. Also, let $(X_v, X_u), (Y_v, Y_u) \in [k] \times [k]$ be two pairs of random variables which are equal to the colour assignments of the pair of vertices (v, u) in the systems S_1 and S_2, correspondingly.

Let $P_{opt}[(X_v, X_u) \neq (Y_v, Y_u)]$ be the probability to have disagreement in the optimal coupling of (X_v, X_u) with (Y_v, Y_u). Let also $P^v_{opt}[X_v \neq Y_v]$ be the probability to have disagreement in the optimal coupling of X_v with Y_v. Let $P^u_{opt}[X_u \neq Y_u]$ be the probability to have disagreement in the optimal coupling of X_u with Y_u.

The coupling lemma (in [2]) suggests that it suffices to show that

$$P_{opt}[(X_v, X_u) \neq (Y_v, Y_u)] \leq P^v_{opt}[X_v \neq Y_v] + P^u_{opt}[X_u \neq Y_u].$$

We are going to give an optimal coupling of (X_v, X_u) with (Y_v, Y_u) which, at the same time, couples optimally X_v with Y_v and X_u with Y_u. Let $\tilde{P}[E]$ be the probability of the event E to occur in the coupling we give, where $E \in [k]^4$.

The coupling consists in partitioning $[0, 1]$ into, at most, k^4 subintervals. Each of these subintervals will correspond to a value of the 4-tuple (X_v, X_u, Y_v, Y_u). The length of the interval that corresponds to $(X_v, X_u, Y_v, Y_u) = (s_1, s_2, s_3, s_4)$ will be equal to $\tilde{P}[(X_v, X_u, Y_v, Y_u) = (s_1, s_2, s_3, s_4)]$. Also, we consider the random variable U that is distributed uniformly in the unit interval $[0, 1]$. Depending on which subinterval of $[0, 1]$ the r.v. U belongs, we assume the corresponding value for the four random variables.

Note that the length of the subintervals should be taken such that

$$\tilde{P}[X_v = s] = Pr[X_v = s] \qquad \forall s \in [k].$$

The analogous should hold for the marginal distributions of X_u, Y_v and Y_u.

We expect to exist subintervals of $[0,1]$ of four kinds. The first kind corresponds to those subintervals for which it holds $X_v = Y_v$ and $X_u = Y_u$. The second one corresponds to those that $X_v = Y_v$ and $X_u \neq Y_u$. The third corresponds to those that $X_v \neq Y_v$ and $X_u = Y_u$. And the last one, corresponds to those that it holds $X_v \neq Y_v$ and $X_u \neq Y_u$.

Consider, first, the subintervals of $[0,1]$, named $l_{s,t}$, for $s,t \in [k]$. The length of the interval $l_{s,t}$ is such that

$$|l_{s,t}| = \min\left\{Pr[(X_v, X_u) = (s,t)], Pr[(Y_v, Y_u) = (s,t)]\right\}.$$

The probability for U to belong to any of these intervals is equal to the sum of their lengths, i.e.

$$\tilde{P}[(X_v, X_u) = (Y_v, Y_u)] = \sum_{s,t \in [k]} |l_{s,t}|.$$

These are the subintervals of the first kind. Furthermore, we can see that there cannot be any coupling that achieves smaller probability than $\sum_{s,t\in[k]} |l_{s,t}|$ for the event "$(X_v, X_u) = (Y_v, Y_u)$". Thus, we get that

$$P_{opt}[(X_v, X_u) \neq (Y_v, Y_u)] = \tilde{P}[(X_v, X_u) \neq (Y_v, Y_u)]. \tag{33}$$

Until know we have shown that our coupling is optimal w.r.t. to coupling the pair (X_v, X_u) with (Y_v, Y_u).

Furthermore, if for $s \in [k]$ it holds $Pr[X_v = s], Pr[Y_v = s] > \sum_{t\in[k]} |l_{s,t}|$, then we can have a subinterval of $[0,1]$ which corresponds to the event, "$X_v = Y_v = s$, $X_u \neq Y_u$". Let $l_{s,*}$ be the name of this subinterval. It is direct that it holds

$$|l_{s,*}| = \min\left\{Pr[X_v = s], Pr[X_u = s]\right\} - \sum_{t\in[k]} |l_{s,t}|.$$

Also, we can see that in our coupling it holds that

$$\tilde{P}[X_v = Y_u] = \sum_{s\in[k]} \min\left\{Pr[X_v = s], Pr[X_u = s]\right\}.$$

Furthermore, we can see that there cannot be any coupling that achieves greater probability than the r.h.s. of the above equality for the event "$X_v = Y_v$". This implies that

$$P^v_{opt}[X_v \neq Y_v] = \tilde{P}[X_v \neq Y_v] \tag{34}$$

Similarly, if for $t \in [k]$ it holds $Pr[X_u = t], Pr[Y_u = t] > \sum_{s\in[k]} |l_{s,t}|$, then we can have a subinterval of $[0,1]$ which correspond to the event, "$X_u = Y_u = t, X_v \neq Y_v$". Let $l_{*,t}$ be this subinterval. It direct that it holds

$$|l_{*,t}| = \min\left\{Pr[X_v = t], Pr[X_u = t]\right\} - \sum_{s\in[k]} |l_{s,t}|.$$

Also, we can see that in our coupling it holds that

$$\tilde{P}[X_u = Y_u] = \sum_{t\in[k]} \min\left\{Pr[X_u = t], Pr[X_u = t]\right\}.$$

Furthermore, we can see that there cannot be any coupling that achieves greater probability than the r.h.s. of the above equality for the event "$X_u = Y_u$". This implies that

$$P^u_{opt}[X_u \neq Y_u] = \tilde{P}[X_u \neq Y_u] \tag{35}$$

The proposition follows by (33), (34) and (35) and, also, by noting that

$$\tilde{P}[(X_v, X_u) \neq (Y_v, Y_u)] \leq \tilde{P}[X_v \neq X_u)] + \tilde{P}[Y_v \neq Y_u]$$

$\diamond$

A.5.4 Proof of Lemma 2.1

Consider the sequence of subgraphs $\mathcal{G}(G) = G_0, \ldots, G_r$, where $r = |E|$ and G_0 is empty. Consider, also, the following telescopic relation

$$|\Omega(G,k)| = |\Omega(G_0,k)| \cdot \prod_{i=0}^{r-1} \frac{|\Omega(G_{i+1},k)|}{|\Omega(G_i,k)|} = k^n \cdot \prod_{i=0}^{r-1} \frac{|\Omega(G_{i+1},k)|}{|\Omega(G_i,k)|}. \tag{36}$$

Lemma will follow by showing that

$$Pr[X_i(u_i) \neq X_i(v_i)] = \frac{|\Omega(G_{i+1},k)|}{|\Omega(G_i,k)|}$$

The above relation clearly holds by noting the following: The set of k-colourings of G_{i+1} is the same as the subset of k-colourings of G_i that contains all the colourings that assign v_i and u_i different colours. Lemma follows.

A.5.5 Proof of Lemma 3.1.

$$\begin{aligned}
||\mu(\cdot|\sigma_x) - \mu(\cdot)||_\Lambda &= \frac{1}{2} \sum_{\sigma_\Lambda \in [k]^\Lambda} |\mu(\sigma_\Lambda|\sigma_x) - \mu(\sigma_\Lambda)| \\
&= \frac{k}{2} \mu(\sigma_x) \sum_{\sigma_\Lambda \in [k]^\Lambda} |\mu(\sigma_\Lambda|\sigma_x) - \mu(\sigma_\Lambda)| \\
&= \frac{k}{2} \sum_{\sigma_\Lambda \in [k]^\Lambda} \mu(\sigma_\Lambda) |\mu(\sigma_x|\sigma_\Lambda) - \mu(\sigma_x)| \\
&\leq \frac{k}{2} \sum_{\sigma_\Lambda \in [k]^\Lambda} \mu(\sigma_\Lambda) \sum_{\tau_x \in [k]} |\mu(\tau_x|\sigma_\Lambda) - \mu(\tau_x)| \\
&\leq k \sum_{\sigma_\Lambda \in [k]^\Lambda} \mu(\sigma_\Lambda) ||\mu(\cdot|\sigma_\Lambda) - \mu(\cdot)||_x.
\end{aligned}$$

Noting that it holds

$$||\mu(\cdot|\sigma_x) - \mu(\cdot|\tau_x)||_\Lambda \leq ||\mu(\cdot|\sigma_x) - \mu(\cdot)||_\Lambda + ||\mu(\cdot) - \mu(\cdot|\tau_x)||_\Lambda$$

we get the second relation of the lemma. For the first inequality we have the following:

$$\begin{aligned}
\sum_{A\in[k]^{\Lambda}} \mu(A)\cdot||\mu(\cdot|A)-\mu(\cdot)||_x &= \frac{1}{2}\cdot\sum_{A\in[k]^{\Lambda}}\sum_{\xi_x\in[k]^x}\mu(A)\cdot|\mu(\xi_x|A)-\mu(\xi_x)| \\
&= \frac{1}{2}\cdot\sum_{\xi_x\in[k]^x}\sum_{A\in[k]^{\Lambda}}\mu(\xi_x)\cdot|\mu(A|\xi_x)-\mu(A)| \\
&= \sum_{\xi_x\in[k]^x}\mu(\xi_x)\cdot\frac{1}{2}\left(\sum_{A\in[k]^{\Lambda}}|\mu(A|\xi_x)-\mu(A)|\right) \\
&= \sum_{\xi_x\in[k]^x}\mu(\xi_x)\cdot||\mu(\cdot|\xi_x)-\mu(\cdot)||_{\Lambda} \\
&\leq \sum_{\xi_x\in[k]^x}\mu(\xi_x)\cdot\max_{\sigma_x,\tau_x\in[k]^x}||\mu(\cdot|\sigma_x)-\mu(\cdot|\tau_x)||_{\Lambda} \\
&= \max_{\sigma_x,\tau_x\in[k]^x}||\mu(\cdot|\sigma_x)-\mu(\cdot|\tau_x)||_{\Lambda}
\end{aligned}$$